

PS Claim 1: Page 82-85; 159pp; Japanese.

XX This sequence represents the protease from *Thermococcus celer* DSM-2476
 CC CC This sequence is a protease of the invention. The proteases of the
 CC CC invention have extremely high thermal stability. The proteases can be
 CC CC used as research reagents, and industrially in the food, drug and
 CC CC chemical industries.

XX Sequence 659 AA.

Query Match 100.0%; Score 3437; DB 18; Length 659;
 Best Local Similarity 100.0%; Pred. No. 2.1e-203;
 Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRGLAVLALVLLAGTALAAPVKPVYRNNAVQKNYGLLTPGLFKKVMNNQV 60
 DB 1 MKRGLAVLALVLLAGTALAAPVKPVYRNNAVQKNYGLLTPGLFKKVMNNQV 60
 QY 61 DTVMFGSGDRDRAVKVRLMGAQVKYSKIIPAVAVKIKARDLLIAGMIDTGYFGNT 120
 DB 61 DTVMFGSGDRDRAVKVRLMGAQVKYSKIIPAVAVKIKARDLLIAGMIDTGYFGNT 120
 QY 121 RVSGIKFIQEDYKVOVDATSVSGIGADTVNWSLGYDGGVVAIVDTGIDANHPDLK 130
 DB 121 RVSGIKFIQEDYKVOVDATSVSGIGADTVNWSLGYDGGVVAIVDTGIDANHPDLK 130
 QY 181 VIGWYDAVNGRSTPYDDGCHTGHVAGTGVNSQYIGVAPGAKLVGKVLGADSGS 240
 DB 181 VIGWYDAVNGRSTPYDDGCHTGHVAGTGVNSQYIGVAPGAKLVGKVLGADSGS 240
 QY 241 VSTIIAGVDMVYQNDKYGIRVNLISGSSQSDGTSLSQAVNNAGDGIYVCAAGNS 300
 DB 241 VSTIIAGVDMVYQNDKYGIRVNLISGSSQSDGTSLSQAVNNAGDGIYVCAAGNS 300
 QY 301 GPNTYTVGSPAAAKVITVYVAVDNDNIASFSSRGPTADGRKPEVAPGVDIAPRAG 360
 DB 301 GPNTYTVGSPAAAKVITVYVAVDNDNIASFSSRGPTADGRKPEVAPGVDIAPRAG 360
 QY 361 TSMGTPINDYTKASGTSNATPHYSVGVALILQAHPSWTPDKVKTALITADIVAPKEIA 420
 DB 361 TSMGTPINDYTKASGTSNATPHYSVGVALILQAHPSWTPDKVKTALITADIVAPKEIA 420
 QY 421 DIAYGAGRVNWKYKAYDDNATFTTGSVADKGSATHFDVSGATFTATLYWDTCSSDI 480
 DB 421 DIAYGAGRVNWKYKAYDDNATFTTGSVADKGSATHFDVSGATFTATLYWDTCSSDI 480
 QY 481 DLXLYDPNGNEVDYSYATYGFYKGVYNTAGTWTWKVYVYKGAANTQVDVSDGSL 540
 DB 481 DLXLYDPNGNEVDYSYATYGFYKGVYNTAGTWTWKVYVYKGAANTQVDVSDGSL 540
 QY 541 SGGGNPNPNPNPTPTTDTOTFTGVSNDYNDTSDTFTNNSGATKITGDLTFTSYND 600
 DB 541 SGGGNPNPNPNPTPTTDTOTFTGVSNDYNDTSDTFTNNSGATKITGDLTFTSYND 600
 QY 601 LDLYDPNGNLVDRSTSSNSYHVEYANPAGTFTFLVAYSTYGHADYQLKAVYIG 659
 DB 601 LDLYDPNGNLVDRSTSSNSYHVEYANPAGTFTFLVAYSTYGHADYQLKAVYIG 659

RESULT 2
 AAW94840 standard; Protein; 659 AA.
 XX
 AC AAW94840;
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE W09856926 Seq ID 12.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin.
 XX
 OS additive; drug; washing agent; foodstuff; chemical synthesis.
 XX
 OS *Thermococcus celer*.

XX W09856926-A1.

XX 17-DEC-1998.

XX 64-JUN-1998; 98WO-JP02455

XX 10-JUN-1997; 97JP-0151565

XX (TAKI) TAKARA SHUZO CO LTD.

XX Asada K, Kato I, Morishita M, Shimojo T, Takakura H;

XX WPI; 1999-080907/07.

XX N-PSDB; AAX05926.

XX Recombinant hyperthermostable protease from *Pyrococcus furiosus*.
 PT and gene encoding it, for large scale production of the protease for
 PT industrial use.

XX Disclosure: Page 55-58; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
 CC at 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-His-Gly-Asp-Pro, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus strairi*) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable
 CC for industrial use, can be used as an additive for drugs, washing agents
 CC and foodstuffs and for chemical synthesis.

XX Sequence 659 AA;

Query Match 100.0%; Score 3437; DB 20; Length 659;

Best Local Similarity 100.0%; Pred. No. 2.1e-203;

Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRGLAVLALVLLAGTALAAPVKPVYRNNAVQKNYGLLTPGLFKKVMNNQV 60
 DB 1 MKRGLAVLALVLLAGTALAAPVKPVYRNNAVQKNYGLLTPGLFKKVMNNQV 60
 QY 61 DTVMFGSGDRDRAVKVRLMGAQVKYSKIIPAVAVKIKARDLLIAGMIDTGYFGNT 120
 DB 61 DTVMFGSGDRDRAVKVRLMGAQVKYSKIIPAVAVKIKARDLLIAGMIDTGYFGNT 120
 QY 121 RVSGIKFIQEDYKVOVDATSVSGIGADTVNWSLGYDGGVVAIVDTGIDANHPDLK 180
 DB 121 RVSGIKFIQEDYKVOVDATSVSGIGADTVNWSLGYDGGVVAIVDTGIDANHPDLK 180
 QY 181 VIGWYDAVNGRSTPYDDGCHTGHVAGTGVNSQYIGVAPGAKLVGKVLGADSGS 240
 DB 181 VIGWYDAVNGRSTPYDDGCHTGHVAGTGVNSQYIGVAPGAKLVGKVLGADSGS 240
 QY 241 VSTIIAGVDMVYQNDKYGIRVNLISGSSQSDGTSLSQAVNNAGDGIYVCAAGNS 300
 DB 241 VSTIIAGVDMVYQNDKYGIRVNLISGSSQSDGTSLSQAVNNAGDGIYVCAAGNS 300
 QY 301 GPNTYTVGSPAAAKVITVYVAVDNDNIASFSSRGPTADGRKPEVAPGVDIAPRAG 360
 DB 301 GPNTYTVGSPAAAKVITVYVAVDNDNIASFSSRGPTADGRKPEVAPGVDIAPRAG 360
 QY 361 TSMGTPINDYTKASGTSNATPHYSVGVALILQAHPSWTPDKVKTALITADIVAPKEIA 420
 DB 361 TSMGTPINDYTKASGTSNATPHYSVGVALILQAHPSWTPDKVKTALITADIVAPKEIA 420
 QY 421 DIAYGAGRVNWKYKAYDDNATFTTGSVADKGSATHFDVSGATFTATLYWDTCSSDI 480
 DB 421 DIAYGAGRVNWKYKAYDDNATFTTGSVADKGSATHFDVSGATFTATLYWDTCSSDI 480

QY 481 DLILYDPNGNEVDYSYTAYGFEKGYINPTAGTWTVKVYSGKAANYQVDVYDGSLSQ 540
 DB 481 DLILYDPNGNEVDYSYTAYGFEKGYINPTAGTWTVKVYSGKAANYQVDVYDGSLSQ 540
 QY 541 SGGGNPNPNPNPTPTTDTOTFTGSDVNDYDTSOTFTMNVSATKIGDITFTSYND 600
 DB 541 SGGGNPNPNPNPTPTTDTOTFTGSDVNDYDTSOTFTMNVSATKIGDITFTSYND 600
 QY 601 LDLYLDPNGLVDRSTSSNYSYHVEYANPAPGTWTFVLVAYSTYGHADYQLKAVYVG 659
 DB 601 LDLYLDPNGLVDRSTSSNYSYHVEYANPAPGTWTFVLVAYSTYGHADYQLKAVYVG 659

RESULT 3

AAW24123
 ID AAW24123 standard; Protein: 659 AA.

AC AAW24123;

DT 20-APR-1998 (first entry)

DE Protease.

KW Protease; research reagent; thermal stability.

OS Synthetic.

XX WO9721823-A1.

PN 19-JUN-1997.

PD 07-NOV-1996; 96WO-JP03253.

PF 12-DEC-1995; 95JP-0323285.

PR (TAKI) TAKARA SHUZO CO LTD.

PI Asada K, Kato I, Mitta M, Morishita M, Takakura H;

PI Tsunashawa S, Yamamoto K;

DR WPI: 1997-332794/30.

DR N-PSDB: AAT85669.

XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are
 useful industrially and as research reagents

PS Claim 9; Page 92-95; 159pp; Japanese.

XX This sequence is a protease of the invention. The proteases of the
 CC invention have extremely high thermal stability. The proteases can be
 CC used as research reagents, and industrially in the food, drug and
 CC chemical industries.

SQ Sequence 659 AA;

Query Match 84.8%; Score 2914; DB 18; Length 659;
 Best Local Similarity 83.7%; Pred. No. 3e-171;
 Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps

QY 1 MKRLGAVLALVLCVLLAGTALAAPKPV--YRNNAVQKNYGLTLPGLKVKVQRMWQ 58

DB 1 MKGLKALILVILGLVGVSAANAEKKVQVRN---VEKNYGLTLPGLKVKVQRMWQ 57

QY 59 EVDVTINFGSYGDRORAVYLRKGAQVKKYKIIPAVAVKIKARDLLIAGHIDGKYG 118

DB 58 EISTVIVENHREKELAVRLELGAQVKKYKIIPAVAVKIKARDLLIAGHIDGKYG 113

QY 119 NTRVSGIFIOEDYKVOVDDA-----TSVQIGACTVNSLGYDGSVVAIVDTGIDAN 173

DB 114 KAKLSGVRFIOEDYKVTSAELEGLEDSAAQVMAVYWN-LGYDSGTTIGITDIDGAS 172

QY 174 HPDLKGVICWYDAVNGRSTPYDQCHCTHVGIVAGTGSV-NSQYTGAVPAGKLVGVAV 232

DB 173 HPDLQSKVIGWDFYNGKSYEYDDHGHGTHTVASTAAGTGAASNGYKGMAPGAKLAGIKV 232
 QY 233 LGADGSGSVSTIIAGDKVYQNKDKYGRVINKLSGSSQSDGTDSLSOAVNNANDAGIV 292
 DB 233 LGADGSGSVSTIIAGDKVYQNKDKYGRVINKLSGSSQSDGTDSLSOAVNNANDAGIV 292
 QY 293 VCVAAGNSGPNNTYTVGSPAAASKVITVGAOSNDNIASFSSRGPTADGRLKPEYVAPGVD 352
 DB 293 VCVAAGNSGPNNTYTVGSPAAASKVITVGAOSNDNIASFSSRGPTADGRLKPEYVAPGVD 352
 QY 353 IIAPRASGTSNGTPTINDYTKASGTSMTATPHVSGVAGLILQAHPSNTPDKVKYKALISTAD 412
 DB 353 IIAPRASGTSNGTPTINDYTKASGTSMTATPHVSGVAGLILQAHPSNTPDKVKYKALISTAD 412
 QY 413 IVAPKEIADIAYGAGRVYKAIKYDDYAKLTFTGSAVCKGATHTEDVSGATEVTATLY 472
 DB 413 IVAPKEIADIAYGAGRVYKAIKYDDYAKLTFTGSAVCKGATHTEDVSGATEVTATLY 472
 QY 473 WDTGSSDIDLVLDPNGNVDYSYTAYGFEKGYINPTAGTWTVKVYSGKAANYQVDV 532
 DB 473 WDTGSSDIDLVLDPNGNVDYSYTAYGFEKGYINPTAGTWTVKVYSGKAANYQVDV 532
 QY 533 VSDGSLGSGGGNPNPNPTPTTDTOTFTGSDVNDYDTSOTFTMNVSATKIGDITFTSYND 592
 DB 533 VSDGSLGSGGGNPNPNPTPTTDTOTFTGSDVNDYDTSOTFTMNVSATKIGDITFTSYND 592
 QY 593 TPDTSYNDLVLDPNGNVDYSYTAYGFEKGYINPTAGTWTVKVYSGKAANYQVDV 652
 DB 593 TPDTSYNDLVLDPNGNVDYSYTAYGFEKGYINPTAGTWTVKVYSGKAANYQVDV 652
 QY 653 KAVVYVG 659
 DB 653 KAVVYVG 659

RESULT 4

AAW24129
 ID AAW24129 standard; Protein: 654 AA.

AC AAW24129;

DT 20-APR-1998 (first entry)

DE Pyrococcus furiosus protease pfus.

KW Protease; research reagent; thermal stability; pyrococcus furiosus.

OS Pyrococcus furiosus DSM-3638.

PN WO9721823-A1.

PD 19-JUN-1997.

PF 07-NOV-1996; 96WO-JP03253.

PR 12-DEC-1995; 95JP-0323285.

PA (TAKI) TAKARA SHUZO CO LTD

PI Asada K, Kato I, Mitta M, Morishita M, Takakura H;

PI Tsunashawa S, Yamamoto K;

DR WPI: 1997-332794/30.

DR N-PSDB: AAT85669.

XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are
 useful industrially and as research reagents

PS Disclosure; Page 125-128; 159pp; Japanese.

XX This sequence represents the protease from Pyrococcus furiosus DSM-3638.

CS-09-841-553-1.rag

Wed Nov 6 14:29:27 2002

CC This sequence is a protease of the invention. The proteases of the
CC invention have extremely high thermal stability. The proteases can be
CC used as research reagents, and industrially in the food, drug and
CC chemical industries.
XX
SQ

Query Match 72.38; Score 2483.5; DB 18; Length 654;
Best Local Similarity 72.18; Pred. No. 8.9e-145;
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

QY 1 MKRLGAVLVLLVGLAGTALAAPKPV--VRNNAVQOKNYGLTPGLFKKVKVRNMNO 58
DB 1 MKGLKALILVILVGLVGVSAAPKKEQVRN---VEKNYGLTLPGLFKKVKVRNMNO 57
QY 59 EVDVIVFGSYGDRDRKAVKRLMGAQVYSYKIIPAVAVKTKAROLLIIAGNIDYVGG 118
DB 58 EISTVIVFENHREKEIAVRVLEKAKVRYVHIIPAIADLKVRDLVLSGL--TG--G 113
QY 119 NTRVSGIKFIOEDYKVOVDDA-----TSVQIGADTVNNSLGVGVVAVYDTGIDAN 173
DB 114 KAKLSGVRFIOEDYKVTSAELEGDESAQVMAIYVNN-LCYDGSGLTIGIITGIDAS 172
QY 174 HPDLGKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSOYIGVAPKLVGVK 232
DB 173 HPDLQKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSOYIGVAPKLVGVK 232
QY 233 LGADGSGSVSTIIAGVYVONKDKYIRVNLISLGSQSSDGTSLSOAVNNADAGIV 292
DB 233 LGADGSGSVSTIIAGVYVONKDKYIRVNLISLGSQSSDGTSLSOAVNNADAGIV 292
QY 293 VCVVAGNSGPNNTYVGSPPAAKSVITVGAVDNDNIASFSSRGPTADGRKLPKVPVAPGN 352
DB 293 VCVVAGNSGPNNTYVGSPPAAKSVITVGAVDNDNIASFSSRGPTADGRKLPKVPVAPGN 352
QY 353 IIAPRASGTSMTGPPINNTKATKATSMATPHVSGVGLIIOAHPSWTPDKVKLTALIED 412
DB 353 IIAARASGTSMTGPPINNTKATKATSMATPHVSGVGLIIOAHPSWTPDKVKLTALIED 412
QY 413 IVPKDEIADIAYGAGRTVYKA FYDDYAKLTGTSVADKCSATHFDVSGATFVATLY 472
DB 413 IVPKDEIADIAYGAGRTVYKA FYDDYAKLTGTSVADKCSATHFDVSGATFVATLY 472
QY 473 WDTGSSDILYLDPNNGNVDYSYATYGFERYGYNPTAGTWTVKVYSYGAANYQDV 532
DB 473 WDNAGSGLDLYLDPNNGNVDYSYATYGFERYGYNPTAGTWTVKVYSYGAANYQDV 532
QY 533 YDGSLSQSGGPNPNPNPTPTDTOTFTGSDVNDYDTSDFMTNNSGATKITGDL 592
DB 533 YDGSLSQSGGPNPNPNPTPTDTOTFTGSDVNDYDTSDFMTNNSGATKITGDL 592
QY 593 PFTSYNDLQLYDNPNGNLDVRSSTNSYEHVEYANPOTWTFVLYAYSTYGHADYOL 652
DB 593 PFTSYNDLQLYDNPNGNLDVRSSTNSYEHVEYANPOTWTFVLYAYSTYGHADYOL 652
QY 653 KAVNYG 659
DB 648 TARKVYG 654

RESULT 5
AAW94841
ID AAW94841 standard; Protein; 654 AA.
XX
AC AAW94841;
XX
DT 06-MAY-1999 (first entry)
XX
DE Hyperthermostable protease.
XX
KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
XX
KW additive; drug; washing agent; foodstuff; chemical synthesis.
XX

OS Pyrococcus furiosus.

XX W095856926-A1.
XX PN 17-DEC-1998
XX PD 04-JUN-1998; 98WO-JP02465.
XX PF 10-JUN-1997; 97JP-0151969.
XX PR (TAKI) TAKARA SHUZO CO LTD.
XX PA Asada K, Kato I, Morishita M, Shimojo T, Takakura H;
XX PI WPI; 1999-080907/07.
XX DR N-PSDB; AAX05929.
XX DR Recombinant hyperthermostable protease from Pyrococcus furiosus -
XX PT and gene encoding it, for large scale production of the protease for
XX PT industrial use.
XX XX Disclosure; Page 60-63; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
XX thermophilic bacterium (especially Pyrococcus furiosus). The protease has
XX working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
XX (optimum 6-8), and retains more than 90% of its activity after 8 hours
XX at 95 deg.C. The invention also provides gene sequences encoding a
XX polypeptide of formula S-G-Ala-Gly-Asn-PRO, where SIG is a signal
XX peptide from subtilisin, and PRO is the above protease. Host cells
XX (especially Bacillus strains) transformed with vectors comprising the
XX genes are used for the recombinant production of the protease. The
XX hyperthermostable protease which can be prepared in quantity suitable
XX for industrial use, can be used as an additive for drugs, washing agents
XX and foodstuffs and for chemical synthesis.

XX Sequence 654 AA:

Query Match 72.38; Score 2403.5; DB 20; Length 654;
Best Local Similarity 72.18; Pred. No. 8.9e-145;
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;
QY 1 MKRLGAVLVLLVGLAGTALAAPKPV--VRNNAVQOKNYGLTPGLFKKVKVRNMNO 58
DB 1 MKGLKALILVILVGLVGVSAAPKKEQVRN---VEKNYGLTLPGLFKKVKVRNMNO 57
QY 59 EVDVIVFGSYGDRDRKAVKRLMGAQVYSYKIIPAVAVKTKAROLLIIAGNIDYVGG 118
DB 58 EISTVIVFENHREKEIAVRVLEKAKVRYVHIIPAIADLKVRDLVLSGL--TG--G 113
QY 119 NTRVSGIKFIOEDYKVOVDDA-----TSVQIGADTVNNSLGVGVVAVYDTGIDAN 173
DB 114 KAKLSGVRFIOEDYKVTSAELEGDESAQVMAIYVNN-LCYDGSGLTIGIITGIDAS 172
QY 174 HPDLGKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSOYIGVAPKLVGVK 232
DB 173 HPDLQKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSV-NSOYIGVAPKLVGVK 232
QY 233 LGADGSGSVSTIIAGVYVONKDKYIRVNLISLGSQSSDGTSLSOAVNNADAGIV 292
DB 233 LGADGSGSVSTIIAGVYVONKDKYIRVNLISLGSQSSDGTSLSOAVNNADAGIV 292
QY 293 VCVVAGNSGPNNTYVGSPPAAKSVITVGAVDNDNIASFSSRGPTADGRKLPKVPVAPGN 352
DB 293 VCVVAGNSGPNNTYVGSPPAAKSVITVGAVDNDNIASFSSRGPTADGRKLPKVPVAPGN 352
QY 353 IIAPRASGTSMTGPPINNTKATKATSMATPHVSGVGLIIOAHPSWTPDKVKLTALIED 412
DB 353 IIAARASGTSMTGPPINNTKATKATSMATPHVSGVGLIIOAHPSWTPDKVKLTALIED 412
QY 413 IVPKDEIADIAYGAGRTVYKA FYDDYAKLTGTSVADKCSATHFDVSGATFVATLY 472
DB 413 IVPKDEIADIAYGAGRTVYKA FYDDYAKLTGTSVADKCSATHFDVSGATFVATLY 472

Qy 141 SVS QIGADTVVNSLGYDGSVVVAIVDTGIDANHPDLKKGICWIDAVNCRSTPYDDQGH 200

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
 CC at 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable
 CC for industrial use, can be used as an additive for drugs, washing agents
 CC and foodstuffs and for chemical synthesis.
 XX
 SQ Sequence 522 AA;
 Query Match 62.2%; Score 2138.5; DB 20; Length 522;
 Best Local Similarity 78.3%; Pred. No. 1.1e-123;
 Matches 407; Conservative 36; Mismatches 70; Indels 7; Gaps 3;
 QY 141 SVSOGADTVNLSGYDGGVVAIVDTGIDANHPDLKGVIGWYDAVNGRSTPYDDQGH 200
 DB 9 SAAQVMATYVNN-LGYDGGSGITIIIDTGDASHEDLQGRVIGWDFVNGRSYPYDDHGH 67
 QY 201 GTHVAGIVAGTGSV-NSOYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVYONKDKYG 259
 DB 68 GTHVASTAAGTGAASNGKYGAPGAKLAGIKVLGADGSGSTIITIKGVEMAYDNKDKYG 127
 QY 260 IRVINLSLGSQSDGTDLSQAVNNAWDAGIVVCVAAGNSGPNMTYTGSPAAASKVITY 319
 DB 128 IKVINLSLGSQSDGTDLSQAVNNAWDAGLVVVAAGNSGPNMTYTGSPAAASKVITY 187
 QY 320 GAVDNDNIASFSGRGTADGRKPEVAPGVVDIIAPRASGTSMTGTPINDYTKASGTSN 379
 DB 188 GAVDKYDVTISFSGRGTADGRKPEVAPGVVDIIAPRASGTSMTGTPINDYTKASGTSN 247
 QY 380 ATPHVSYGVALILOAHPSTPDKVKYKTALLETADIIVAPKEIADIYAGGRVNYKAKYDD 439
 DB 248 ATPHVAGIAALLQAHPSWTPDKVKYKTALLETADIIVAPKEIADIYAGGRVNYKAKYDD 307
 QY 440 YAKLFTGTVADKGSATHTFDVNSGATFVATLYMDTGSDDILYDPPNGNEVDYSYAT 499
 DB 308 YAKLVFTGYVANKGSQTHQFVNSGATFVATLYMDTGSDDILYDPPNGNEVDYSYAT 367
 QY 500 YGFEKVGYNPTAGTWTWVYKGAANYQVDVSDGLSOSGGGNPNPNPTPTPTD 559
 DB 368 YGFEKVGYNPTAGTWTWVYKGAANYQVDVSDGLSOSGGGNPNPNPTPTPTD 422
 QY 560 TQFTGTVNDYMDSDTFTMNNSGATKITGDLTFTSYNDLILYDPPNGNLDVRSFS 619
 DB 423 AKTFQXSDHYYYDRSDFTFTVNSGATKITGDLVFTSYNDLILYDPPNGNLDVRSFS 482
 QY 620 NSYEHVEYANPAGTWTWVYKGAANYQVDVSDGLSOSGGGNPNPNPTPTPTD 659
 DB 483 NSYEHVEYANPAGTWTWVYKGAANYQVDVSDGLSOSGGGNPNPNPTPTPTD 522
 RESULT 8
 AAW94836
 ID AAW94836 standard; Protein; 412 AA.
 XX
 AC AAW94836;
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE Hyperthermostable protease fragment.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis.
 XX
 OS *Pyrococcus furiosus*.
 XX
 PN W09856926-A1.

XX 17-DEC-1998.
 PD
 XX
 PF 04-JUN-1998; 98WO-JP02453.
 XX
 PR 10-JUN-1997; 97JP-0151969.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Asada K, Kato I, Morishita M, Shimomo T, Takakura H;
 XX
 XX WPI: 1999-080907/07.
 DR N-PSDB; AAX05920.
 XX
 PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* -
 PT and gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Claim 2; Page 35-37; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
 CC at 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable
 CC for industrial use, can be used as an additive for drugs, washing agents
 CC and foodstuffs and for chemical synthesis.
 XX
 SQ Sequence 412 AA;
 Query Match 49.7%; Score 1707; DB 20; Length 412;
 Best Local Similarity 80.6%; Pred. No. 2.9e-97;
 Matches 325; Conservative 29; Mismatches 47; Indels 2; Gaps 2;
 QY 141 SVSOGADTVNLSGYDGGVVAIVDTGIDANHPDLKGVIGWYDAVNGRSTPYDDQGH 200
 DB 9 SAAQVMATYVNN-LGYDGGSGITIIIDTGDASHEDLQGRVIGWDFVNGRSYPYDDHGH 67
 QY 201 GTHVAGIVAGTGSV-NSOYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVYONKDKYG 259
 DB 68 GTHVASTAAGTGAASNGKYGAPGAKLAGIKVLGADGSGSTIITIKGVEMAYDNKDKYG 127
 QY 260 IRVINLSLGSQSDGTDLSQAVNNAWDAGIVVCVAAGNSGPNMTYTGSPAAASKVITY 319
 DB 128 IKVINLSLGSQSDGTDLSQAVNNAWDAGLVVVAAGNSGPNMTYTGSPAAASKVITY 187
 QY 320 GAVDNDNIASFSGRGTADGRKPEVAPGVVDIIAPRASGTSMTGTPINDYTKASGTSN 379
 DB 188 GAVDKYDVTISFSGRGTADGRKPEVAPGVVDIIAPRASGTSMTGTPINDYTKASGTSN 247
 QY 380 ATPHVSYGVALILOAHPSTPDKVKYKTALLETADIIVAPKEIADIYAGGRVNYKAKYDD 439
 DB 248 ATPHVAGIAALLQAHPSWTPDKVKYKTALLETADIIVAPKEIADIYAGGRVNYKAKYDD 307
 QY 440 YAKLFTGTVADKGSATHTFDVNSGATFVATLYMDTGSDDILYDPPNGNEVDYSYAT 499
 DB 308 YAKLVFTGYVANKGSQTHQFVNSGATFVATLYMDTGSDDILYDPPNGNEVDYSYAT 367
 QY 500 YGFEKVGYNPTAGTWTWVYKGAANYQVDVSDGLSOSGGGNPNPNPTPTPTD 542
 DB 368 YGFEKVGYNPTAGTWTWVYKGAANYQVDVSDGLSOSGGGNPNPNPTPTPTD 410
 RESULT 9
 AAR87009
 ID AAR87009 standard; Protein; 237 AA.
 XX
 AC AAR87009;

XX 10-MAY-1996 (first entry)
 XX Hyperthermostable protease.
 XX Protease; hyperthermostable; thermostability.
 XX Pyrococcus furiosus.
 XX Key Location/Qualifiers
 XX Misc-difference 196 /note= "unspecified amino acid"
 XX W09534645-A1
 XX 21-DEC-1995
 XX 05-JUN-1995; 95WO-JP01095.
 XX 26-JUL-1994; 94JP-0173912.
 XX 13-JUN-1994; 94JP-0130236.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Asada K, Kato I, Mitsu M, Morishita M, Tsunashima S;
 XX Yamamoto K;
 XX WPI: 1996-049674/05.
 XX N-PSDB: AAT08131.
 XX Pyrococcus furiosus hyperthermostable protease gene - useful for
 XX recombinant prodn. of hyperthermostable protease
 XX Example 3; Page 70-71; 85pp; Japanese.
 XX The invention relates to the hyperthermostable protease of
 XX Pyrococcus furiosus and its prodn. as a recombinant protein
 XX in transformants using a vector carrying the protease gene
 XX (AAT08131). A genomic DNA sequence of the invention is given in
 XX AAT08131 and its encoded protein in AAT87009.
 XX Sequence 237 AA;

Query Match 33.6%; Score 1154; DB 17; Length 237;
 Best Local Similarity 97.4%; Pred. No. 1.5e-63;
 Matches 224; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 OY 176 DLKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKGLA 235
 DB 1 DLKGVIGWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVKGLA 60
 OY 236 DCGSGVSTIAGVDVWVQNKDYGIRVNLGLSGSSQSDGDTLSQVNNAMDAGIVCV 295
 DB 61 DCGSGVSTIAGVDVWVQNKDYGIRVNLGLSGSSQSDGDTLSQVNNAMDAGIVCV 120
 OY 295 AAGNSGPNVTYVSPAAASKVITVGAVDSDNDNIASFSSRGPTADGRLKPEVYVAPGVDTIA 355
 DB 121 AAGNSGPNVTYVSPAAASKVITVGAVDSDNDNIASFSSRGPTADGRLKPEVYVAPGVDTIA 180
 OY 356 PRASGTSNGTPINDYTKASGTSMTATPHVSGVGLIILQARPSNTPDKVKT 405
 DB 181 PRASGTSNGTPINDYKNKSGSSMDTPHVSQVGLIILQARPSNTPDKVKT 230

RESULT 10
 AAR87011
 ID AAR87011 standard; Peptide; 188 AA.
 XX AAR87011;
 XX 10-MAY-1996 (first entry)
 XX Peptide sequence.

XX Protease; hyperthermostable; thermostability.
 XX Pyrococcus furiosus.
 XX W09534645-A1.
 XX 21-DEC-1995.
 XX 05-JUN-1995; 95WO-JP01095.
 XX 26-JUL-1994; 94JP-0173912.
 XX 13-JUN-1994; 94JP-0130236.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Asada K, Kato I, Mitsu M, Morishita M, Tsunashima S;
 XX Yamamoto K;
 XX WPI: 1996-049674/05.
 XX Pyrococcus furiosus hyperthermostable protease gene - useful for
 XX recombinant prodn. of hyperthermostable protease
 XX Example 4; Page 66-67; 85pp; Japanese.
 XX The invention relates to a gene (AAT08131) that codes for a
 XX hyperthermostable protease (AAR87007) of Pyrococcus furiosus.
 XX 2 DNA sequences (AAT08131-1) are provided encoding peptides
 XX (AAR87010-11).
 XX Sequence 188 AA;
 Query Match 27.1%; Score 935; DB 17; Length 188;
 Best Local Similarity 98.4%; Pred. No. 3.3e-50;
 Matches 184; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 200 HGTTHAGIVAGTGSVNSQYIGVAPGAKLVGVKGLADGSGSVSTIAGVDVWVQNKDYG 259
 DB 2 HGTTHAGIVAGTGSVNSQYIGVAPGAKLVGVKGLADGSGSVSTIAGVDVWVQNKDYG 61
 OY 260 IRVNLGLSGSSQSDGDTLSQVNNAMDAGIVCVVAAGNSGPNVTYVSPAAASKVITV 319
 DB 62 IRVNLGLSGSSQSDGDTLSQVNNAMDAGIVCVVAAGNSGPNVTYVSPAAASKVITV 121
 OY 320 GAVSDNDNIASFSSRGPTADGRLKPEVYVAPGVDTIAPRASGTSMTGTPINDYTKASGTS 379
 DB 122 GAVSDNDNIASFSSRGPTADGRLKPEVYVAPGVDTIAPRASGTSMTGTPINDYTKASGTS 181
 OY 380 ATPHVSQ 386
 DB 182 ATPHVTG 188

RESULT 11
 AAR24126
 ID AAR24126 standard; Protein; 188 AA.
 XX AAR24126;
 XX 20-APR-1998 (first entry)
 XX Thermococcus protease fragment.
 XX Protease; research reagent; thermal stability; thermococcus celer.
 XX Thermococcus celer DSM-24126.
 XX W09721823-A1.
 XX 19-JUN-1997.
 XX 07-NOV-1996; 96WO-JP03233.

```

XX PR 12-DEC-1995; 95JP-0323285.
XX PA (TAKI) TAKARA SHUZO CO LTD.
XX PI Asada K, Kato I, Mita M, Morishita M, Takakura H;
XX PI Tsunashima S, Yamamoto K;
XX DK WPI: 1997-332794/30.
XX DR N-PSDB; AAT85676.
XX PS
XX PT Protease(s) and genes encoding them obtained from Thermococcus and
XX PT Pyrococcus strains - have extremely high thermal stability and are
XX PT useful industrially and as research reagents
XX PS Disclosure; Page 110-112; 159pp; Japanese.
XX CC This sequence represents a fragment of the protease from Thermococcus
XX CC celer DSM-2476 (see AAW24121 for full length sequence). This sequence is
XX CC a fragment of the protease of the invention. The proteases of the
XX CC invention have extremely high thermal stability. The proteases can be
XX CC used as research reagents, and industrially in the food, drug and
XX CC chemical industries.
XX SQ Sequence 188 AA;
Query Match 27.2%; Score 935; DB 18; Length 188;
Best Local Similarity 98.4%; Pred. No. 3.3e-50;
Matches 184; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 200 HGTWAGVACVNSYIGVAPGAKLVGVGLGADGGSVYIIAGVYVYVQNKDKY 215
DB 2 HGTWAGVAGTGGVNSYIGVAPGAKLVGVGLGADGGSVYIIAGVYVYVQNKDKY 61
QY 260 IRVNLSSGSSDSDTSLQAVNNAWDAGIVVCAAGNSGPNYTVGSPAAASKVITV 319
DB 62 IRVNLSSGSSDSDGSLQAVNNAWDAGIVVCAAGNSGPNYTVGSPAAASKVITV 121
QY 370 GAVDSNDNTASFSSRGPTADGRLAPEVVPVGVVDIAPRASGTSMTGPIINDYTKASGISM 379
DB 122 GAVDSNDNTASFSSRGPTADGRLAPEVVPVGVVDIAPRASGTSMTGPIINDYTKASGISM 181
QY 380 ATPHVS 386
DB 182 ATPHVG 188
RESULT 12
AAW13667
ID AAW13667 standard; Protein; 734 AA.
XX AC AAW13667;
XX DT 06-OCT-1997 (first entry)
XX DE Streptomyces viridosporus dhpa gene product.
XX KW asymmetric hydrolase; dhpa; 4-substituted-1,4-dihydropyridine;
XX KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
XX KW cardiovascular; treatment; hypertension; ischaemic heart disease
XX OS Streptomyces viridosporus.
XX FH Key Location/Qualifiers
XX FT Peptide 205..724
XX FT /note= "fragment of dhpa protein; see AAW13667"
XX PN W09705243-A1.
XX XX 13-FEB-1997.
XX PD 30-JUL-1996; 96WO-JP02147.
XX XX

```

```

PR 29-FEB-1996; 96JP-0067479.
PR 31-JUL-1995; 95JP-0212973.
XX PA (SAOC) MERCIAN CORP.
XX PI Arisawa A, Dobashi K, Tashiki K, Matsufuji M, Nakashima T;
XX PI Tsuruta T, Yoshio K;
XX DK WPI: 1997-145682/13.
XX DR N-PSDB; AAT61454.
XX PS
XX PT Asymmetric hydrolase gene derived from Streptomyces viridosporus -
XX PT acts on 4-substituted-1,4-dihydropyridine derivatives to produce
XX PT chiral derivatives useful for synthesis of cardiovascular drugs
XX PS Claim 3; Page 49-55; 78pp; Japanese.
XX CC This sequence is an asymmetric hydrolase encoded by the Streptomyces
XX CC viridosporus dhpa gene. The enzyme acts on 4-substituted-1,4-
XX CC dihydropyridine derivatives. The enzyme allows the efficient conversion
XX CC of 4-substituted-1,4-dihydropyridine esters to chiral partially
XX CC hydrolysed derivatives, for use in the synthesis of cardiovascular drugs
XX CC suitable for the treatment of e.g. hypertension and ischaemic heart
XX CC disease.
XX SQ Sequence 734 AA;
Query Match 19.3%; Score 665; DB 18; Length 734;
Best Local Similarity 31.2%; Pred. No. 7.1e-33;
Matches 216; Conservative 34; Mismatches 269; Indels 124; Gaps 26;
QY 22 LAAPKPVVRNNAVQKNYGLFPGLEKKVQRMNWNQVDTVIMFGSYGDRDRANKVRL 81
DB 97 LPDAARLVASCKLDRFLDITELG--KAATRSQKQGLKIV---CYQGAARAAAEVR 151
QY 82 MGAQVYKVIIPAVAKIKARALLIAGMIDTGVFGNTRVSGIKFTQDYKVQVDDATS 141
DB 152 EAGELARTLTLNADAVRTPTPT---ASELNDVYNGDRTASGIAHWLQGVRAALDTS 208
QY 142 VSGIGADTVNNSLGVDSGVV---VDTGIDANHPDLKKVIGWYDAYNGRSTP--YDDOG 199
DB 209 VGGIGAPKAW-SAGYDCKGVKIAVLDTGVDTSHPDLKRVTA---SKNFTAAPGADKVG 264
QY 200 HGTWAGVAGTGSVN-SOYIGVAPGAKLVGVGLGADGGSVYIIAGVYVYVQNKDY 258
DB 265 HGTWASIAAGTGAQSKKYGVARGAAILNGKVLDDSGDGGIILAGMEWAAQ---- 320
QY 259 GIRVINLSGSSQSDGDSLSQAVNN-ANDAGIVVCAAGNSGPNYTVGSPAAASKVI 317
DB 321 GADVNMNSLGGMDTPE--TDPLEAAVKLSAERGLVDFALANEGEPE--SIGSPGSDAAL 377
QY 318 TVGAVDSNDNTASFSSRGPTADGRLAPEVVPVGVVDIAPRASGTSMTGPIIND--YYTK 373
DB 378 TVGAVDDKDLADFSSTGRLGDCAIKPDVTPAGVDITAAAEAGNDIGQEVGEGPAGYMT 437
QY 374 ASGTSNATPHNSVGCALLIQHP-SMTPTDKVKLTALIEADIAPKEIADIAYGAGRVNVYK 433
DB 438 ISGTSNATPHVAGAAALNQOHP-TWSEALKGAL--TGSTAGCK-YTPFEQSGSGRIQADK 494
QY 434 AAK-----YDDYAKLTPTGSVADKGSNTH 457
DB 495 ALQQTVIADPVSFQVCPVPHNTDEPVTQLTYRNLGQDVTLKLTSTATDPKGAAPA 554
QY 458 TFDVSSGATFVATLYWDGSS--HOLLYDPNGNEVDYSYATYIGFEKVGYYNPATGWTV 517
DB 555 GFFTLGATTVTPA---G--SASVDNTADTRLGCTVDGAYSAYVWATGGQVTRTAAAVQR 611
QY 518 KVSYS-----KGAENQVDVDSGLSSQSGGNNPNPNPNPTPTDTCFT--- 563
DB 612 EVESYDVTVRIGRDKPPT--T--TDLTGAGLQSGRGY-----APATDTATLRLP 662
QY 564 TGS-VNDYNDTSDTFM-----NVNSGATKITGDLTFT--SYNDLDLYLPNGNLY 613

```

Db 663 KGTLYDSNIAKDFGLKGGIDMLVQPKLSVTKDT-TLTLDARTTKAADITVDPDK----- 717
 QY 614 DRSTSSNSYEHVEYANPAPGWTFLVAYSTYG 646
 Db 718 -----AKPLSAT---ICVYDTAG 733

RESULT 13
 AA013668
 ID AA013668 standard; Protein: 823 AA.
 AC AA013668;
 DT 06-OCT-1997 (first entry)
 DE DhpA-mel chimeric protein.
 KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease.
 OS Chimeric Streptomyces viridosporus.
 OS Chimeric Streptomyces antibioticus.
 PH Key Location/Qualifiers
 FT Protein 1..734
 FT Peptide /label= DhpA_protein_product
 FT Peptide 205..724
 FT Peptide /note= "see AA013666"
 FT Peptide 735..823
 FT /note= "melanin (partial sequence)"
 W09705243-A1.
 PD 13-FEB-1997.
 XX 30-JUL-1996; 96WO-JP02147.
 XX 29-FEB-1996; 96JP-0067478.
 PR 31-JUL-1995; 95JP-0212975.
 PA (SACC) MERCIAN CORP.
 PI Arisawa A, Dobashi K, Iashiki K, Matsufuji M, Nakashima T;
 PI Tsuruta T, Yoshiohka T;
 DR WPI; 1997-145682/13.
 DR N-PSDB; AAT61455.
 PT Asymmetric hydrolase gene derived from Streptomyces viridosporus -
 PT acts on 4-substituted-1,4-dihydropyridine derivatives to produce
 PT chiral derivatives useful for synthesis of cardiovascular drugs
 PS Claim 5; Page 37-43; 78pp; Japanese.
 CC This sequence is a fusion protein comprising Streptomyces viridosporus
 CC dhpA gene product (an asymmetric hydrolase) which acts on 4-substituted-
 CC 1,4-dihydropyridine derivatives, and melanin from S. antibioticus. The
 CC dhpA enzyme allows the efficient conversion of 4-substituted-1,4-
 CC dihydropyridine esters to chiral partially hydrolysed derivatives,
 CC for use in the synthesis of cardiovascular drugs suitable for the
 CC treatment of e.g. hypertension and ischaemic heart disease.
 SQ Sequence 823 AA;

Query Match 19.3%; Score 665; DB 18; Length 823;
 Best Local Similarity 31.2%; Pred. No. 8.2e-33;
 Matches 216; Conservative 84; Mismatches 269; Indels 124; Gaps 26;
 QY 22 LAAPKPVYRNAYQORNGULTPGLFKKVKORMMNOEVDYTFMGSGDGRDAVKVRL 81
 Db 97 LPADAARLVASGKLDORLEDTBLG--KAATRSQKQGLKIV---GYGGAARAAKACVR 151

QY 82 MGAQVYSYKIIPAVAVKIVARDLLLIACKIDTGYFQNTSRVSGIKFIQBYKVOVDATS 141
 Db 152 EAGELRRLTSLNADAVYTRPHED---ASELMDAVTNGDRTASIAHVMIDGVRAALOTS 208
 QY 142 VSOIGADTVNNSLGYDGSYVAIVDTGIDANHPDLKGVIGHYDAVNGRSTP---YDDOG 199
 Db 209 VQIGAPKAW-SAGYDGGVKIATVLTGDTSHPLKGRVTA---SKNFTAPGAGCKVG 264
 QY 200 HGTHVAGIVAGTGSVN-SGVIGVAPGAKLVGVLGADGSGSVSTIIAGVDMVYQNNKDY 258
 Db 265 HGTHVASIAAGTGAQSGKGYKGVAPGAAILNGAVLDDSGFGDSDGILAGNEWAAQ---- 320
 QY 259 GIRVINLSGSSQSDGTSLQOAVNN-AMDAGTVVCVYAGNSGPNVTYVGSPPAAASKVI 317
 Db 321 GADVYNSLGGCHDTPF-TQPLAAVVKLSAEKGVLEIAAGNEGPE--SIGSPGSADAAL 377
 QY 318 TVGAVDSNDNTASFSKGP-TADERLKPVEVAPGVDTIAPRASGTSMTCPIND---YYTK 373
 Db 378 TVGAYDDKDKLAQFSSTGPRLDGAIKPVDVTAQVDTITAASAGNDIGQEVGEGPACWT 437
 QY 374 ASGTSNATPHVSGVGLIILQAHESWTPOKVKTALETADIAPKEIADIAYGAGRVNYTK 433
 Db 438 ISGTSNATPHVAGAAALGAGCHPMTSALKGAL--TGSTKGGK-YTPFEGSGSRIOADK 494
 QY 434 ATK-----YDDYAKLTFGTGSVADKGSATH 457
 Db 495 ALQQTIVADPVSFVGVQVWPHPTDDEPVTKQLYRNLGTQDVTLLKLTSTATDPKGAAPA 554
 QY 458 TFDVSGATFTVATLYMDTCSIDILYDPNGNEVDYKSYTAYVGFENKGVYNNPTACTWT 517
 Db 555 GFETLGATTVTPA---GGVASVDNTADTRLGTVDGAYSAIVVATGGGQTVRTAAVOR 611
 QY 518 KVSYSY-----KGAHVQVDVVSQSLSSQSGGNPNPNPNPTPTDTDTOTF--- 563
 Db 612 EVESYDVTVRHIGROGKPTKELIDDLIGVAGLGSRGYV-----APATDTATRLP 662
 QY 564 TGS-VNDYDHTSDTFM-----NVNSGATKITGDLTDT-SYNDLQLYDPNGLV 613
 Db 663 KGTLYDSNIAKDFGLKGGIDMLVQPKLSVTKDT-TLTLDARTTKAADITVDPDK----- 717
 QY 614 DRSTSSNSYEHVEYANPAPGWTFLVAYSTYG 646
 Db 718 -----AKPLSAT---ICVYDTAG 733

RESULT 14
 AA013666
 ID AA013666 standard; Protein: 820 AA.
 AC AA013666;
 DT 06-OCT-1997 (first entry)
 DE Fragment of dhpA gene product.
 KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease.
 OS Streptomyces viridosporus.
 PH Key Location/Qualifiers
 FT MISC-difference 29
 FT MISC-difference 238
 FT MISC-difference 238 /note= "asp29 as indicated in the specification"
 FT MISC-difference 238 /note= "asp232 as indicated in the specification"
 PN W09705243-A1.
 XX 13-FEB-1997.
 XX 30-JUL-1996; 96WO-JP02147.

MS-09-841-553-1.rag

Wed Nov 6 14:29:27 2002

PR 29-FEB-1996; 96JP-0067478.
 PR 31-JUL-1995; 95JP-0212975.
 XX (SAOC) MERCIAN CORP.
 PI Arisawa A, Dobashi K, Isshiki K, Matsufuji M, Nakashima T;
 PI Tsuruta T, Yoshioka T;
 XX WPI; 1997-145682/13.
 DR N-85DB; AAT61454-55.
 XX Asymmetric hydrolase gene derived from Streptomyces viridosporus -
 PT acts on 4-substituted-1,4-di-hydro:pyridine derivatives to produce
 PT chiral derivatives useful for synthesis of cardiovascular drugs
 XX Claim 1; Page 60-63; 78pp; Japanese.
 PS This sequence is a fragment of an asymmetric hydrolase which acts on
 CC 4-substituted-1,4-dihydropyridine derivatives. The DNA sequence encoding
 CC the hydrolase (phdA) was isolated from the chromosomal DNA of
 CC Streptomyces viridosporus. The enzyme allows the efficient conversion
 CC of 4-substituted-1,4-dihydropyridine esters to chiral partially
 CC hydrolysed derivatives, for use in the synthesis of cardiovascular drugs
 CC suitable for the treatment of e.g. hypertension and ischaemic heart
 CC disease.
 XX Sequence 520 AA; Length 520;
 Query Match 18.8%; Score 647.5; DB 18; Indels 85; Gaps 18;
 Best Local Similarity 35.9%; Pred. No. 5.5e-32;
 Matches 179; Conservative 58; Mismatches 177;
 QY 140 TSVSQTGADTVNWSLGDGSGVVAIVDGDANHPDLKGVWDAVNGRSTP--YSD 197
 DB 3 TSVGQIGAPKAW-SAGVDGKGVIAVLDTGDTSHPLKGRYA---SKNFTAPGAGK 58
 QY 198 QGHTGHVAGTGSVN-SQI-GVAPGAKLVGVVGLADGSGSVSTIAGVDVWVQND-256
 DB 59 VGHTGHVASTAGTGAQSKGKYGVAPGAAILNCKVLDGSGDGGSLAGHENAQAQ--116
 QY 257 KYGIRVIMLSGSSSDGTSLSQAVNN-ANDAGIVVCAVAGNSPNTYVGSPPAASR 315
 DB 117 --GADVNNSLAGMDTP-TDPLERAVDKLSAEKGVLFATAAGNEGPE--SIGSPGSDA 172
 QY 316 VITVGVADSDNINAFSSRGP-TADGRKPEVAVPGVDIAPRASGTSMTGPIIND---Y 371
 DB 172 ALTGVAVDDKDLADESFGRLGDAIKEDVTAAGVDTAASAEKNDIGQVGGPAGY 332
 QY 372 TRASGTSMATPHVSGVALLIQAHPSTPKYKTALETADIVAPKEIADIAYGAGRVV 332
 DB 232 MTISGTSMATPHVAGAAALLKQHPDWTSAELKAGAL--TGSTGKG-YTPFEGSGRLQA 240
 QY 432 YKAIK-----YDYAKLFTSGVADKGA 455
 DB 289 DKALQOTVIADPVSVSGVQVQNPHTDDEPVTKQTYRNLGTODVTLKLTATIDPKGAA 349
 QY 456 THFFVSGATVFTATLWOTGSSDIDLVDPNNGNEVYSYATYGFERYGYNPTAGTW 515
 DB 349 PAGFTTLAGTTVTPA---GGSAVDHMTADTFLGGTVGAGSAYVYVATGGGTAVTAAY 405
 QY 516 TVKAVSY-----KGAANYQVDVSDGSLSSGSGGNPNPNPNTPTTDTQTF- 563
 DB 406 QREVESYDVTVRHIGROGKPTTEHLTDLIGVAGLSGRCVG-----APATDTATLR 453
 QY 564 --TGS-VNDYNDTSDTFTM 579
 DB 457 LFKGTLYVDSWLANDFGTL 475
 RESULT 15
 AAB81180
 ID AAB81180 standard; Protein: 1079 AA.
 XX

AAB81180;
 13-JUL-2001 (first entry)
 Transglutaminase related protein SEQ ID 39.
 Coryneform bacteria; transglutaminase; food processing.
 Streptomyces albobogiseolus.
 WC200123591-A1.
 05-APR-2001.
 29-SEP-2000; 2000WO-JP06780.
 30-SEP-1999; 99JP-0280098.
 28-JUN-2000; 2000JP-0194043.
 (AJIN) AJINOMOTO CO INC.
 Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;
 WPI; 2001-266172/27.
 Efficient secretory production of foreign proteins e.g.
 transglutaminase employing transformant coryneform bacterium, simply on
 industrial scale with direct recovery for use in food processing and
 pharmaceutical industry
 Disclosure; Page 113-120; 151pp; Japanese.
 This invention relates to a process for the production of a foreign
 secretory protein through the construction of a recombinant coryneform
 bacterium. The coryneform bacterium is transformed with an expression
 construct in which DNA encoding a target foreign protein is fused to a
 signal peptide of a coryneform bacterium. The recombinant bacterium is
 cultured in a medium containing a carbon source and a nitrogen source.
 The recombinant bacterium produces the foreign protein. The foreign
 protein is purified from the culture medium. The foreign protein is
 useful in the food processing and pharmaceutical industries. The present
 sequence represents a transglutaminase related protein. Transglutaminases
 are used in examples illustrating the method of the invention.

Query Match 18.5%; Score 636; DB 22; Length 1079;
 Best Local Similarity 32.2%; Pred. No. 6.9e-31;
 Matches 201; Conservative 72; Mismatches 219; Indels 132; Gaps 26;
 QY 118 GNTVRSGIKFIOED--YKVOVDVATSVSQIGADTVNWSLGDGSGVVAIVDGTIDANHP 175
 DB 153 GDTASGVAVWLDGVKASL--TSVGQIGTPKAMEA-GYDCKGVKIAVLDTGVDATHP 209
 QY 176 DLKGVKTVGWDVAVNGRSTPY--TPOGHGTHVAGTGSVN-SQYIGVAPGAKLVGVK 232
 DB 210 DLKGVTA---SKNFTSAPT--VGHTGHVASTAGTGAQSKGTYKGVAPGAKLVGVK 266
 QY 233 LGADGSGSVSTIAGVDVWVQNP-KYKVICRVIMLSGSSSDGTSLSQAVNN-ANDAGI 291
 DB 267 LDDAGFGDSDGSLAGHENAQAQ-----GADVNNSLAGMDTP-TDPLERAVDKLSAEKGI 321
 QY 292 VVCAVAGNSPNTYVGSVAPKASKVITVGVADSDNINAFSSRGP-TADGRKPEVAVPG 350
 DB 322 LFAAAGNEGPO---SIGSPGSDALTVCAVDKDLADESFGRLGDAIKEDVTAAGVDTA 379
 QY 351 VDIAPRASGTSMTGPIINDY---YTKASGTSMATPHVSGVALLIQAHPSTPKYKTALET 407
 DB 380 VDIATAASAKGNDIAKEVGEFPGAGYMTISGTSMATPHVAGAAALLKQHPDWTSAELK 439
 QY 408 -TETADIVAPKEIADIAYGAGRVVYKAIKY-----YDYAKLFTSGVADKGA 444
 DB 440 TASTKD-----GKVTPEQSGRVVDKAITQTVIAPVSVSGVQVQNPHTDDEPVTKKLT 495

US-09-841-553-1.rai

Wed Nov 6 14:29:28 2002

US-09-445-472-12

Query Match 100.0% Score 3437 DB 4 Length 659;
Best Local Similarity 100.0% Pred. No. 6.7e-237; Indels 0; Gaps 0;
Matches 659; Conservative 0; Mismatches 0;

QY 1 MKRGAVVLAIVLGLAGTALAAPKPVVRNNAVQKNYGLTGLPKFKKVVORNNHVEV 60
DB 1 MKRGAVVLAIVLGLAGTALAAPKPVVRNNAVQKNYGLTGLPKFKKVVORNNHVEV 60
QY 61 DTVINEFSYSGDRRAVVKRLMKAQVYKIKIIPAVAVKIKARDLLIAGMIDTGYGNT 120
DB 61 DTVINEFSYSGDRRAVVKRLMKAQVYKIKIIPAVAVKIKARDLLIAGMIDTGYGNT 120
QY 121 RVSGTKFQEDYKQVDDATSVSQIGADTVNNSLGYDGSVYVAIVDTGIDANHPDLK 180
DB 121 RVSGTKFQEDYKQVDDATSVSQIGADTVNNSLGYDGSVYVAIVDTGIDANHPDLK 180
QY 181 VIGWDVANGSTPYDQGHGTHYAGTVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGS 240
DB 181 VIGWDVANGSTPYDQGHGTHYAGTVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGS 240
QY 241 VSTTIAGVDMVYQNKIKYIRVNLISLGSOSSDGTDSLQAVNNAMDAGIVVYVAAAGNS 300
DB 241 VSTTIAGVDMVYQNKIKYIRVNLISLGSOSSDGTDSLQAVNNAMDAGIVVYVAAAGNS 300
QY 301 GPNTYTVGSPAAASKVITVGAFTSDNDNIASFSSRGPTADGRKPEVAPGVVDIAPRSG 360
DB 301 GPNTYTVGSPAAASKVITVGAFTSDNDNIASFSSRGPTADGRKPEVAPGVVDIAPRSG 360
QY 361 TSNCTPINDYTKASGTSMAKFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
DB 361 TSNCTPINDYTKASGTSMAKFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
QY 421 DIATYAGRVNVAIKYDDYAKLFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
DB 421 DIATYAGRVNVAIKYDDYAKLFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
QY 481 DLVLYDPNGNEVDYSYATYGFYKGVYNYPTAGTWTWVYKSYKGAANYQVVDVSDGSLQ 540
DB 481 DLVLYDPNGNEVDYSYATYGFYKGVYNYPTAGTWTWVYKSYKGAANYQVVDVSDGSLQ 540
QY 541 SGGGNPNPNPNPTPTDTDTFTGVSNDYMDTSDFTMNVNSGATKITGDLTPTSYND 600
DB 541 SGGGNPNPNPNPTPTDTDTFTGVSNDYMDTSDFTMNVNSGATKITGDLTPTSYND 600
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYATYTGWADYQKAVYVG 659
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYATYTGWADYQKAVYVG 659

RESULT 3
US-08-894-818B-5
Sequence 5, Application US/08/894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YANAMOTO, Katsunao
APPLICANT: MIYATA, Masanori
APPLICANT: ASADA, Kiyoko
APPLICANT: TSUNASAWA, Shunmu
APPLICANT: KATO, Ikumoshin
TITLE OF INVENTION: HYPER-EXPRESSIBLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM

US-08-894-818B-1

Query Match 100.0% Score 3437 DB 4 Length 659;
Best Local Similarity 100.0% Pred. No. 6.7e-237; Indels 0; Gaps 0;
Matches 659; Conservative 0; Mismatches 0;

QY 1 MKRGAVVLAIVLGLAGTALAAPKPVVRNNAVQKNYGLTGLPKFKKVVORNNHVEV 60
DB 1 MKRGAVVLAIVLGLAGTALAAPKPVVRNNAVQKNYGLTGLPKFKKVVORNNHVEV 60
QY 61 DTVINEFSYSGDRRAVVKRLMKAQVYKIKIIPAVAVKIKARDLLIAGMIDTGYGNT 120
DB 61 DTVINEFSYSGDRRAVVKRLMKAQVYKIKIIPAVAVKIKARDLLIAGMIDTGYGNT 120
QY 121 RVSGTKFQEDYKQVDDATSVSQIGADTVNNSLGYDGSVYVAIVDTGIDANHPDLK 180
DB 121 RVSGTKFQEDYKQVDDATSVSQIGADTVNNSLGYDGSVYVAIVDTGIDANHPDLK 180
QY 181 VIGWDVANGSTPYDQGHGTHYAGTVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGS 240
DB 181 VIGWDVANGSTPYDQGHGTHYAGTVAGTGSVNSQYIGVAPGAKLVGVKVLGADSGS 240
QY 241 VSTTIAGVDMVYQNKIKYIRVNLISLGSOSSDGTDSLQAVNNAMDAGIVVYVAAAGNS 300
DB 241 VSTTIAGVDMVYQNKIKYIRVNLISLGSOSSDGTDSLQAVNNAMDAGIVVYVAAAGNS 300
QY 301 GPNTYTVGSPAAASKVITVGAFTSDNDNIASFSSRGPTADGRKPEVAPGVVDIAPRSG 360
DB 301 GPNTYTVGSPAAASKVITVGAFTSDNDNIASFSSRGPTADGRKPEVAPGVVDIAPRSG 360
QY 361 TSNCTPINDYTKASGTSMAKFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
DB 361 TSNCTPINDYTKASGTSMAKFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
QY 421 DIATYAGRVNVAIKYDDYAKLFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
DB 421 DIATYAGRVNVAIKYDDYAKLFTGSGVADGSAHTFDVSGATEVATLYMDTSSDI 480
QY 481 DLVLYDPNGNEVDYSYATYGFYKGVYNYPTAGTWTWVYKSYKGAANYQVVDVSDGSLQ 540
DB 481 DLVLYDPNGNEVDYSYATYGFYKGVYNYPTAGTWTWVYKSYKGAANYQVVDVSDGSLQ 540
QY 541 SGGGNPNPNPNPTPTDTDTFTGVSNDYMDTSDFTMNVNSGATKITGDLTPTSYND 600
DB 541 SGGGNPNPNPNPTPTDTDTFTGVSNDYMDTSDFTMNVNSGATKITGDLTPTSYND 600
QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYATYTGWADYQKAVYVG 659
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTFELVYATYTGWADYQKAVYVG 659

RESULT 2
US-09-445-472-12
Sequence 12, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMIZU, Tomoko
APPLICANT: ASADA, Kiyoko
APPLICANT: KATO, Ikumoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPER-EXPRESSIBLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472
CURRENT FILING DATE: 1999-12-06
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 12
LENGTH: 659
TYPE: PRT
ORGANISM: Thermococcus celer

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/894,818B
 FILING DATE: 20-MAY-1998
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA: PCT/JP96/03253
 FILING DATE: 07-NOV-1996
 PRIOR APPLICATION NUMBER: JP 323285/1995
 FILING DATE: 12-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Browdy, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: TAKAKURA-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 659 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-894-818B-5

Query Match 84.8%; Score 2914; DB 4; Length 659;
 Best Local Similarity 83.7%; Pred. No. 1.le-199;
 Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY	1	MRLGCAVLAALVGLVGLAGTALAAAPKPY--YRNNAVQKNYGLTPGLFKKQVPMNQ 58
DB	1	MGLKALILVILVGLVGLVGSAAAPKPKVEQVRN---VEKNYGLTPGLFKKQVPMNQ 57
QY	59	EVDVTFMGSGYGRDRRAVRLKMAQKYSKIIPAVYKIKARDLLIAGMTDTGYG 118
DB	58	EISTVTFVFNHREKREIAVRVLELMGAKRVYVHIIPAADLKVRLVLSGL--TG--G 113
QY	119	NTRVSGIKFTQDYKYVOVDA-----TSVSGIGADTWNLSGLVDCGCVVAIVDTGIDAN 173
DB	114	KAKLSVRFIOEDYKVTVAEGLDGSAAQVATYVNN--LGVDGSGITIGIIDTGDAS 172
QY	174	HPDLGKGVIGHYDAVNGRSTPYDDOCHGTHVAGIVAGTGSV--NSQYIGVAPGAKLVGV 131
DB	173	HPDLGKGVIGHYDAVNGRSTPYDDOCHGTHVAGTGSV--NSQYIGVAPGAKLVGV 131
QY	233	LGADSGSVSTIAGVDMVYVQNKDKYGRVNLNLGSSQSDGTSLSQAVNNANDAGIV 202
DB	233	LGADSGSVSTIAGVDMVYVQNKDKYGRVNLNLGSSQSDGTSLSQAVNNANDAGIV 202
QY	293	VCVAAGNSGPNVTYVGSPPAAASKVITVGAVDSDNINIAFSSRGTADGLKPEVYAPGVD 352
DB	293	VCVAAGNSGPNVTYVGSPPAAASKVITVGAVDSDNINIAFSSRGTADGLKPEVYAPGVD 352
QY	353	IIAPRASGTSMTGPIINDYTKASGTSMTATPHVSGVGLIILQAHPSWTPDKYKTALLETAD 412
DB	353	IIAPRASGTSMTGPIINDYTKASGTSMTATPHVSGVGLIILQAHPSWTPDKYKTALLETAD 412
QY	413	IVAPKETADYAGGRVNYKAIKYDDYAKLTFTGVSADGSAHTFDFVSGATFVATLY 472
DB	413	IVAPKETADYAGGRVNYKAIKYDDYAKLTFTGVSADGSAHTFDFVSGATFVATLY 472
QY	473	WDTGSSDIDLVLDPNGNEVDYSYATYTGFEKVGYYNPTAGTWTVKVYSYKGAANYQVDV 532
DB	473	WDTGSSDIDLVLDPNGNEVDYSYATYTGFEKVGYYNPTAGTWTVKVYSYKGAANYQVDV 532
QY	533	YDSGLSOSGGGNPNPNPNPTPTDTOTFTGVSNDYWDTSFTFNVNSGATKIGDL 592
DB	533	YDSGLSOSGGGNPNPNPNPTPTDTOTFTGVSNDYWDTSFTFNVNSGATKIGDL 592

QY 593 TFDTSYNDLDLYDPNGVGRSTSSNSYEHVEYANPAPGCTWTFVLVYAYSTYGMADYOL 652
 DB 593 TFDTSYNDLDLYDPNGVGRSTSSNSYEHVEYANPAPGCTWTFVLVYAYSTYGMADYOL 652
 QY 653 KAVVYIG 659
 DB 653 KAVVYIG 659

RESULT 4
 US-08-894-818B-35
 : Sequence 35, Application US/08894818B
 : Patent No. 6261822
 : GENERAL INFORMATION:
 : APPLICANT: TAKAKURA, Hikaru
 : APPLICANT: MORISHITA, KIC
 : APPLICANT: YAMAMOTO, Katsuhiko
 : APPLICANT: MITTA, Masahiro
 : APPLICANT: ASADA, Kiyozo
 : APPLICANT: TSUNASAKA, Susumu
 : APPLICANT: KATO, Ikunobu
 : TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 : NUMBER OF SEQUENCES: 42
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Browdy and Nelmark
 : STREET: 419 Seventh Street N.W., Ste. 300
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: United States of America
 : ZIP: 20004
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/894,818B
 : FILING DATE: 20-MAY-1998
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PCT/JP96/03253
 : FILING DATE: 07-NOV-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 323285/1995
 : FILING DATE: 12-DEC-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Browdy, Roger L.
 : REGISTRATION NUMBER: 25,618
 : REFERENCE/DOCKET NUMBER: TAKAKURA-1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 628-5197
 : TELEFAX: (202) 737-3528
 : INFORMATION FOR SEQ ID NO: 35:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 654 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: peptide
 : US-08-894-818B-35

Query Match 72.3%; Score 2483.5; DB 4; Length 654;
 Best Local Similarity 72.1%; Pred. No. 5.le-169;
 Matches 481; Conservative 59; Mismatches 96; Indels 21; Gaps 8;

QY	1	MRLGCAVLAALVGLVGLAGTALAAAPKPY--YRNNAVQKNYGLTPGLFKKQVPMNQ 58
DB	1	MGLKALILVILVGLVGLVGSAAAPKPKVEQVRN---VEKNYGLTPGLFKKQVPMNQ 57
QY	59	EVDVTFMGSGYGRDRRAVRLKMAQKYSKIIPAVYKIKARDLLIAGMTDTGYG 118
DB	58	EISTVTFVFNHREKREIAVRVLELMGAKRVYVHIIPAADLKVRLVLSGL--TG--G 113

119 NTRVSGIKFIOEDYKVVDDA-----TSVQIGADTWNLSLGVDSGVVVAIVDTGIDAN 173
114 KAKLSGVRFIOEDYKVVDSAELEGLDESAQAQVMAIYWN-LGYDGSIGITIGITGIDAS 172
174 HPDLKGVIGWYDAVNGRSTPYDDOCHGTHVAGIVAGTGSV-NSQYIGVAPCAKLVGVK 232
173 HPDLKGVIGWYDAVNGRSTPYDDOCHGTHVAGIVAGTGSV-NSQYIGVAPCAKLVGVK 232
233 LGADSGSVSTIAGVDMVYVONKDKYIRVNLGLSSQSDGTDLSQAVNNAWDAGIV 292
233 LGADSGSVSTIAGVDMVYVONKDKYIRVNLGLSSQSDGTDLSQAVNNAWDAGIV 292
293 VCVAAGNSGPNNTYVGSPPAAAKVITVGAVDSDNDNIASFSSRGPTADGRLKPEVYVAPGV 352
293 VCVAAGNSGPNNTYVGSPPAAAKVITVGAVDSDNDNIASFSSRGPTADGRLKPEVYVAPGV 352
353 IIAARASGTSNGPTINDYTKASGTSMAHPVSGVAGLILQAHPSWTPDKVKALITETAD 412
353 IIAARASGTSNGPTINDYTKASGTSMAHPVSGVAGLILQAHPSWTPDKVKALITETAD 412
413 IVAPKEIADIAGVAGRNVAIKYDDYAKLTFTGSVADKGSATHFVDSGATFVTATLY 472
413 IVAPKEIADIAGVAGRNVAIKYDDYAKLTFTGSVADKGSATHFVDSGATFVTATLY 472
473 WDTGSSDILYLDPNNGNEVDYSYATYGFYKGVYINPTAGITWTVKVSYKGAANYQVDV 532
473 WDTGSSDILYLDPNNGNEVDYSYATYGFYKGVYINPTAGITWTVKVSYKGAANYQVDV 532
533 VSDGSLQSGGGNPNPNPPTPTDTOTFTGVSNDYWDTSFTTNNVNSGATKITGDL 592
533 VSDGSLQSGGGNPNPNPPTPTDTOTFTGVSNDYWDTSFTTNNVNSGATKITGDL 592
593 TFDTSYNDLILYLDPNNGNEVDYSYATYGFYKGVYINPTAGITWTVKVSYKGAANYQVDV 652
593 TFDTSYNDLILYLDPNNGNEVDYSYATYGFYKGVYINPTAGITWTVKVSYKGAANYQVDV 652
653 KAVVYIG 659
648 TAKVYIG 654

RESULT 6
US-08-894-818B-3
Sequence 3, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Kazuhiko
APPLICANT: MITA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

US-09-445-472-16
Sequence 15, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOTO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPER-THERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445.472
CURRENT FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 16
LENGTH: 654
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-445-472-16

Query Match 72.3%; Score 2483.5; DB 4; Length 654;
Best Local Similarity 72.1%; Pred. No. 5.1e-169;
Matches 481; Conservative 69; Mismatches 96; Indels 21; Gaps 8;

QY 1 MKRLGAVVLAIVLVGLLAGTAAAPKVPV--VRNNAVQKNYGLTLPGLPKYQBMWQ 58
DB 1 MKRLGAVVLAIVLVGLLAGTAAAPKVPV--VRNNAVQKNYGLTLPGLPKYQBMWQ 57

RESULT 7
US-09-445-472-4
; Sequence 4, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:

RESULT 8
US-09-445-472-1
; Sequence 1, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, MIO
; APPLICANT: SHIMOTO, Tomoko
; APPLICANT: ASADA, Kiyozo

CURRENT APPLICATION DATA: 33
APPLICATION NUMBER: US/63/750,532
FILING DATE: 13-JUN-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA: PCT/JP95/01095
APPLICATION NUMBER: PCT/JP95/01095
FILING DATE: 05-JUN-1994
PRIORITY APPLICATION DATA: JP 994,130,236
APPLICATION NUMBER: JP 994,130,236
FILING DATE: 13-JUN-1994
PRIORITY APPLICATION DATA: JP 924/173912
APPLICATION NUMBER: JP 924/173912
FILING DATE: 26-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 35,616
REFERENCE/DOCKET NUMBER: MITTA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08/750-532-18

Query Match	33.4%	Score 1148;	DB 1;	Length 237;
Best Local Similarity	97.0%	Pred. No. 1.6e-74;		
Conservative	1;	Mismatches 6;	Indels 0;	Gaps 0;

176	QY	DLKGVKIGWDVANGNGASTPYDQIGTGTHTAGIVAGTGVSSVSSQIIGVAPGKRLVGVKVLGA	60
1	DB	DLKGVKIGWYDAVNGRSTPYDQIGTGTHTAGIVAGTGVSSVSSQIIGVAPGKRLVGVKVLGA	
236	QY	DGSGSVSTIAGVDMVYVONKDKYKRVINLSLASSOSSDGTDSLQAVNNADAGIWCY	295
61	DB	DGSGSVSTIAGVDMVYVONKDKYKRVINLSLASSOSSDGTDSLQAVNNADAGIWCY	120
396	QY	AAGNSGPNITYVGSPPAAASKVITVGAVDSDNDNIASFSSRGPTADGRLAPEVYAPGVDDIIA	355
131	DB	AAGNSGPNITYVGSPPAAASKVITVGAVDSDNDNIASFSSRGPTADGRLAPEVYAPGVDDIIA	180
356	QY	PRASGISMGTPIINDITYKASCTSMATPHVSGVGLILQAHPSTPDKVKT	405
181	DB	PRASGISMGTPIINDITYKASCTSMATPHVSGVGLILQAHPSTPDKVKT	230

01 JUL 1968

RES-09-000-016-4
 Sequence 4, Application US/090007016
 Patent No. 6143541
 GENERAL INFORMATION:
 APPLICANT: AKIRA ARIKAWA, et al.
 TITLE OF INVENTION: GENETIC CODING A PROTEIN HAVING
 TITLE OF INVENTION: HEMOLYSE ACTIVITY FOR 4-
 TITLE OF INVENTION: ITS EXPRESSION PRODUCT
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 STREET: 2033 K Street, N.W., #800
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 1.44 mb
 COMPUTER: IBM compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:

APPLICANT: KATO, Ikunoshin
 TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 FILE REFERENCE: TAKAKURA-6
 CURRENT APPLICATION NUMBER: US/09/445,472
 CURRENT FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: 151969/1997
 PRIOR FILING DATE: 1997-06-10
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 1
 LENGTH: 412
 TYPE: prt
 ORGANISM: Pyrococcus furiosus
 US-09-445-472-1

49.78; Score 1707; DB 4; Length 412;
Query Match 80.6%; Pred. No. 5.3e-114;
Best Local Similarity 29; Mismatches 47; Indels 2; Gaps 2;
Matches 325; Conservative

141	QY	SVYSGADTWNKSLGYDGSQVVAIVTDGTIDANHPOLKRGVLMFLDAYAKG	67
9	DD	SAAQWNAVYVN-LGYDGSGLTIGITDGTIDASHRPLQCKVIGWDFVNGRSTPYDDHGH	67
201	QY	GTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGYDWMYQNKDKYG	359
68	DB	GTHVASIAGTGAASNGYKGMAPGNKLKVLGADGSGSITIIKGVEMAVDNDKYG	377
260	QY	IRVINLSLGSQSSDGTDSLQAVNNANDAGIVVYVCAAGNSGPNYTVGSPAAASKVTV	319
128	DB	IKVINLSLGSQSSDGTDALQAVNANDAGLVYVVAAGNSGPNKYTVLGSPPAAASKVTV	187
320	QY	GAVDSDNDNTASFSSRGPTADGRULKPEVAPGVYDIIAPRASGTSKMGPTINDYTTKASGTSM	379
188	DB	GAVDKVDVITISFSSRGPTDGRULKPEVAPGVNIIAARASGTSKMGPTINDYTTAAPTSM	247
380	QY	ATPHVSGVAGALIIQABPSWTPDKVKYALLETADIYAPKEIADJAYGAGRVNYKAIKYD	439
248	DB	ATPHVAGIAALLQAHPSWTPDKVKYALLETADIYKPEIADJAYGAGRVNAYKAIYDN	307
440	QY	YAKLTPTFGSVADKGAATHTFVQSGATEVATILYDTGSSDIDLVLDPNGREVDYXTAY	495
308	DB	YAKLVFTFGYVANKSGTOHFVQSGAPVATILYDPMNANSDLDLVLDPNGNQVDYXTAY	367
500	QY	YGFEKGVYNYPTAGTHTYKVVYSYKGAANYQDVVDSGLSOSG	542
368	DB	YGFEKGVYNYPTDGTWIIKWSYSGSNYQDVVDSGLSQPG	410

RESULT 9

US-08-750-532-18
Sequence 18, Application US/08750532
Patent No. 5756339
GENERAL INFORMATION:
APPLICANT: MITTA, Nasanori
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MORISHITA, Mio
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street N.W., Suite 300
City: Washington
STATE: D.C. United States of America
COUNTRY:
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

APPLICATION NUMBER: US/09/000.016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 734 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-000-016-4

Query Match 19.34; Score 665; DB 4; Length 734;
Best Local Similarity 31.24; Pred. No. 1.8e-39;
Matches 216; Conservative 84; Mismatches 269; Indels 124; Gaps 26;

QY 22 LAAPVKPVYNNNAQKNGVLLTPGLFKYVQRMNNQEVDTVINFSGYGRDRAVKVRL 81
DB 97 LPDAARLVASGKLDORRFDITELG--KAATRSORGLKLV---GYQAAARAAAEVR 151
QY 82 MGAQVYSYKIIPAVAVKIKARDLLIAGMIDTGFNTRVSGIKFIQEDYKVOVDDAT 141
DB 152 EAGELRRLTSLMADAVRPHED---ASELMDAVTNGDRTASGIAHVWLDGVRRAALDS 208
QY 142 VSOIGADTVNSLGTGSGVVAIVDTGIDANHPDLKVKYGDWDAVNGRSTP--YDDOG 199
DB 209 VGOIGAPKAM--SAGIDGKGVIAVLDITGDTSHPLDKGRVTA---SKNFTAAPGADKVG 254
QY 200 HGTIVAGIVAGTGSVN--SOYTGAPAGAKLVGVKVLGADGSGSVTIIAGVDMVYVQNKDY 259
DB 265 HGTHVSIAGTGAQSKYKGVAPGAAILNKGKVLDDSGDGGDGLAGMENAQAQ---1320
QY 259 GIRVINLSGSSQSDGTSLSQAVNN--ANDAGIVVCAAGNSGPNYTVGSPAASKVI 317
DB 321 GADVNNKSLGGMDTPE--TDPLEAAVOKLSAEKGVLFATAAGNEGE--SIGSPGSADAAL 377
QY 318 TVGAVSDNINIAFSRCP--TADGRLEKPEVWAPGVVDIIAPRASGTSMTPIND---YYTK 373
DB 378 TVCAVDKOKLADFSSTGRLGDKATKPDVTAGVDITAAAEAGNDIGQEVGEGPAGYMT 437
QY 374 ASGTSMTAPVSGVGCALILQHPFSWTPDKVKTALLETADIVAPKEIADIAVCAGRVNYK 433
DB 438 ISGTSMTAPVHVAAGAAALLKQHPDWTSAELKCAL--TGSTKGGK--VTPFQSGSGRIQADK 494
QY 434 AIK-----YDDYAKLTFTGVSADKGSATH 457
DB 495 ALQQTVIADPVSVSGVQVQHPHTDDEPVTKQLTYNLCTQVTLKLTSTATDPKKAAPA 554
QY 458 TPDVSGATVYATLYWDTGSDLDLYLDPNGNEVDVSYTAIYGFKVGYYNPTAGTWT 517
DB 555 GFTLTGNTTVVPA---GGSAVDMTADTRLGCTVCGAYSAYVAVTGGQTVRTAAAVOR 611
QY 518 KVVSY-----KGAANYQDVVSDGSLSSQSGGNPNPNPNPTPTDQTF---563
DB 612 EVESYDVTYRIHGRGKPTTEHLTDLIGYAGLSGRGV-----APATDTATL 662
QY 564 TGS--VNDYWDSTFTN-----NWSGATKLTGDLTDT--SYNDLDLYDPNSM 613
DB 663 KCTYLVDSHIAKDCTGLKGDMLVOPKLSVTKDT--TLTLDARTKAADITVPOPK-----717
QY 614 DRSTSSNYEHVRYANPAGTWTFLVYAISTYG 646

DB 718 -----AKPLSAT---ICGYTYDTAG 733

RESULT 11

US-09-000-016-2
Sequence 2: Application US/090000016
Patent No. 6143541

GENERAL INFORMATION:

APPLICANT: Akira ARISANA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Mendocino, Lind & Ponack, L.L.P.
STREET: 2013 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000.016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 823 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-000-016-2

Query Match 19.34; Score 665; DB 4; Length 823;

Best Local Similarity 31.24; Pred. No. 2.1e-39;

Matches 216; Conservative 84; Mismatches 269; Indels 124; Gaps 26;

QY 22 LAAPVKPVYNNNAQKNGVLLTPGLFKYVQRMNNQEVDTVINFSGYGRDRAVKVRL 81
DB 97 LPDAARLVASGKLDORRFDITELG--KAATRSORGLKLV---GYQAAARAAAEVR 151
QY 82 MGAQVYSYKIIPAVAVKIKARDLLIAGMIDTGFNTRVSGIKFIQEDYKVOVDDAT 141
DB 152 EAGELRRLTSLMADAVRPHED---ASELMDAVTNGDRTASGIAHVWLDGVRRAALDS 208
QY 142 VSOIGADTVNSLGTGSGVVAIVDTGIDANHPDLKVKYGDWDAVNGRSTP--YDDOG 199
DB 209 VGOIGAPKAM--SAGIDGKGVIAVLDITGDTSHPLDKGRVTA---SKNFTAAPGADKVG 254
QY 200 HGTIVAGIVAGTGSVN--SOYTGAPAGAKLVGVKVLGADGSGSVTIIAGVDMVYVQNKDY 259
DB 265 HGTHVSIAGTGAQSKYKGVAPGAAILNKGKVLDDSGDGGDGLAGMENAQAQ---1320
QY 259 GIRVINLSGSSQSDGTSLSQAVNN--ANDAGIVVCAAGNSGPNYTVGSPAASKVI 317
DB 321 GADVNNKSLGGMDTPE--TDPLEAAVOKLSAEKGVLFATAAGNEGE--SIGSPGSADAAL 377

```

US-09-000-016-7

Query Match      18.88, Score 647.5; DB 4; Length 520;
Best Local Similarity 35.9%; Pred. No. 28-38; Indels 85; Gaps
Matches 179; Conservative 54; Mismatches 177;

QY 140 TSVSQIGADTVNSLGYDGSQVYAIVDTGIDANHPOLCKYIGNYDAVNGRSTP--YDD 197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 TSVQIGADPKAN--SAGYDEKGVKIAVLDTGYDTSHPDLKGRVTA---SKNFATAAPGAGDK 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 198 QGHETHVAGTVAGTSVN--SOYIGVAPGAKLVGVKVLGADSGSUSTIAGDVVYQNKD 256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 VGHGTHVASTAAGTGAQSGKRY:KGVAPGAAIILUNGKVLDDSGDDSGIILAHENWRAAQ-- 116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 257 KYGIRVNLISLGSOSSDGTGTSLSQAVNN--AMDAGIVVCVAAGNSPNTYTVGSPAASK 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 --GADYNNKSLGMDTPB-TD?LENAVDLKSAEKGVLFAIAGNEGPE--SIGSPGSA 171
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 316 VITGVADSDNDNTASFSSRGP--TAGDRLKPEVPAVGVDITAPASGTSNGTPIIND--YY 371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 ALTGVADDDKDLADESSFGPRLGCGAIKFPDTPAGVDITPAASBGNDIGQEVGSPAGY 231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 372 TKASGTSMTNPVSHVSGVALLIQAHPSPTDPKVTALITETADIVAPKEIADIAYGAGRVNV 431
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 NITGTSNATPHVAGAAALCKQHPDNTSALKGAL--TGSTKCGK-YTFPFGQSGRIQA 288
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 432 YKAIK-----YDDYAKLFTTGSVADKGA 348
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 DKALQOQTVIADPVSVSFGVQNP--DDDEPVTKOLYRNLCQDVTLKLTSTATDFPKGAA 515
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 456 THPFDVSGATEVTATLYWPTGE:SDILVLYDPNGNEVDYSTAYIGFERYGYYNNTAGTW 515
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 PAGFFTLGATVTVPA---CGSASVDMTADTRLGGVFDGASVYVAATGGQOTVTRAAV 408
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 516 TVKWSY-----KQANYQVDVWSDGSLSGSGGNNPNPNPNPTPTTDTQTF- 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 QREVEYDVTVNHILGRDCK?PTEHLTDLIGVAGLGSRGYG-----APATDTATLR 456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 564 --TGS-VNDYDWTDSYTFM 579
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 LPKGTLYVDWSWLAKDFGTI 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13
US-08-685-774-2
Sequence 2, Application US/08#65774
Patent No. 5712147
GENERAL INFORMATION:
APPLICANT: Shih, Jason C. H.
APPLICANT: Lin, Xiang
APPLICANT: Miller, Eric S.
TITLE OF INVENTION: DNA CODING BACILLUS LICHENIFORMIS
TITLE OF INVENTION: DNA CODING BACILLUS LICHENIFORMIS
TITLE OF INVENTION: DNA CODING BACILLUS LICHENIFORMIS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenneth D. Gibbey
STREET: Post Office Drawer 34009
CITY: Charlotte
STATE: NO. 5712147th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 03/05/685,774
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/CO.250.028
FILING DATE: 27-MAY-1994

ATTORNEY/AGENT INFORMATION:
 NAME: Sibley, Kenneth D.
 REGISTRATION NUMBER: 31,665
 REFERENCE/DOCKET NUMBER: 5051-260
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 420-2200
 TELEFAX: (919) 881-3175
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 379 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORGANISM: Bacillus licheniformis
 STRAIN: PWD-1
 US-08-685-774-2

Query Match 16.38; Score 560.5; DB 1; Length 379;
 Best Local Similarity 35.08; Pred. No. 2e-32;
 Matches 150; Conservative 53; Mismatches 133; Indels 49; Gaps 14;

QY 57 NOEVDVTFNGSYGDRDRAYK--VLRMGAGVQKYSYKIIPAVAVKIKARDLLIAGMIDT 114
 DB 35 NVEKDYIVGPKS-GVKTSVKQVIGSGKVDKQFRLINAAKAKLDKALKEVKNDDPV 93
 QY 115 GYFGNTRVSGIKFIQEDYKVOVDDAT---SVSQIGADTVNNSLGYDGGVVAIVDTGID 171
 DB 94 AY-----VEEDVHAHALAQTVPGYIPLIKADKV-QAOGFGANVAVLDGTGID 141
 QY 172 ANHPDLKGVIGWDAVNGRSTPYDDQGHGTHVAGTGVNSQYIGVAPCAKLVGVK 231
 DB 142 ASHPDL--NVYGASPFAGEAYNTDNGHGHGTHVAGTVAALDNTTG-VLGVAPSVSLYAVK 198
 QY 232 VLGADGSGSVSTIAGVDMVYQNKDYGLRVINLSLSSGSSDGTDSLQAVNNADAGI 291
 DB 199 VLNSSGSGSYSGIVSGIENATTN---GMDVINNSLG---GASGSTATKQAVDNAYARGV 251
 QY 292 VYCVAAAGNSGP--WTVYVGSFAASKVITVGVADSDNDNIASFSSRCPTADGRKPEYVAP 349
 DB 252 VYVAAAGNSGSGNTWITGYPAKYSVIAVGVADSDNSNRASFSSVG-----AELEVMAP 305
 QY 350 GVDIIAPRASGTSNGTPIINDYYTKASCTSMATPHVSGVGCALILQAHPSWTDPKVKYKALIE 409
 DB 306 G-----AGVYSTYPTNTYAT-LNGTSNVSPHVAGAAALILSKHENLSASQVNRLLSS 356
 QY 410 TADIVAPKEIADIATGACRVNRYKA 434
 DB 357 TATYLG-----SSPYGKGLINVEAA 377

RESULT 14
 US-07-923-260A-4
 Sequence 4; Application US/07923260A
 Patent No. 5719021
 GENERAL INFORMATION:
 APPLICANT: Incuys, Masayori
 TITLE OF INVENTION: PROTEIN ACTIVATION
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Gerard J. Weiser
 STREET: 230 South Fifteenth Street, Suite 500
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/923,260A
 FILING DATE: 31-JUL-1992
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Weiser, Gerard
 REGISTRATION NUMBER: 19,763
 REFERENCE/DOCKET NUMBER: 377,5638P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-875-8383
 TELEFAX: 215-875-8394
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 350 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Bacillus licheniformis
 FEATURE:
 NAME/KEY: Region
 LOCATION: 1..76
 OTHER INFORMATION: /note= "The region from 1 to 76 is
 OTHER INFORMATION: a propeptide."
 US-07-923-260A-4

Query Match 16.24; Score 557.5; DB 1; Length 350;
 Best Local Similarity 37.58; Pred. No. 2.9e-32;
 Matches 146; Conservative 56; Mismatches 134; Indels 49; Gaps 13;

QY 57 NOEVDVTFNGSYGDRDRAYK--VLRMGAGVQKYSYKIIPAVAVKIKARDLLIAGMIDT 114
 DB 6 NVEKDYIVGPKS-GVKTSVKQVIGSGKVDKQFRLINAAKAKLDKALKEVKNDDPV 64
 QY 115 GYFGNTRVSGIKFIQEDYKVOVDDAT---SVSQIGADTVNNSLGYDGGVVAIVDTGID 171
 DB 65 AY-----VEEDVHAHALAQTVPGYIPLIKADKV-QAOGFGANVAVLDGTGID 112
 QY 172 ANHPDLKGVIGWDAVNGRSTPYDDQGHGTHVAGTGVNSQYIGVAPCAKLVGVK 231
 DB 113 ASHPDL--NVYGASPFAGEAYNTDNGHGHGTHVAGTVAALDNTTG-VLGVAPSVSLYAVK 169
 QY 232 VLGADGSGSVSTIAGVDMVYQNKDYGLRVINLSLSSGSSDGTDSLQAVNNADAGI 291
 DB 170 VLNSSGSGSYSGIVSGIENATTN---GMDVINNSLG---GASGSTATKQAVDNAYARGV 222
 QY 292 VYCVAAAGNSGP--WTVYVGSFAASKVITVGVADSDNDNIASFSSRCPTADGRKPEYVAP 349
 DB 223 VYVAAAGNSGSGNTWITGYPAKYSVIAVGVADSDNSNRASFSSVG-----AELEVMAP 276
 QY 350 GVDIIAPRASGTSNGTPIINDYYTKASCTSMATPHVSGVGCALILQAHPSWTDPKVKYKALIE 409
 DB 277 GACVYSTYPTST-----VYVLTGTSMAHPVACMAALILSKIPNLSASQVNRLLSS 327
 QY 410 TADIVAPKEIADIATGACRVNRYKA 434
 DB 328 TATYLG-----SSPYGKGLINVEAA 348

RESULT 15
 US-08-845-295A-1
 Sequence 1; Application US/08845295A
 Patent No. 5817490
 GENERAL INFORMATION:
 APPLICANT: Hubbs, John C.
 TITLE OF INVENTION: Enzymatic Process for the Manufacture of
 TITLE OF INVENTION: Ascorbic Acid, 2-Keto-L-Gulonic Acid, and Esters of
 TITLE OF INVENTION: 2-Keto-L-Gulonic Acid
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Eastman Chemical Company
 STREET: P.O. Box 511

us-09-841-553-1.lai

Wed Nov 6 14:29:28 2002

CITY: Kingsport
STATE: Tennessee
COUNTRY: USA
ZIP: 37662-5075
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,295A
FILING DATE: 25-April-97
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 60/017,879
APPLICATION NUMBER: 17-MAY-1996
FILING DATE: 17-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Cheryl J. Tubach
REGISTRATION NUMBER: 38,346
REFERENCE/DOCKET NUMBER: 70432
TELECOMMUNICATION INFORMATION:
TELEPHONE: 423-229-5189
TELEFAX: 423-229-1239
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-08-845-295A-1

Query Match 16.2% Score 557.5; DB 2; Length 379;
Best Local Similarity 37.9%; Pred. No. 3.3e-32;
Matches 146; Conservative 56; Mismatches 134; Indels 49; Gaps 13;

QY 57 NQEVDTVIMFGSYGRDRVAVK--VLRKAGQVYKSYKIIPAVAVKIKARDLLLIAGHMTT 114
DB 35 NVEKDVIVGFKS-GVKTSYKKDIIRESGGKVDKQFRIINAkakLDKEALKEVKNDPOV 93
QY 115 GYFGNTRVSGINFIQEDYKVVDDAT---SVSGIGADTVVNSLIGDGGVYVAIVDTGID 171
DB 94 AY-----VEEDHVAHALAQTPYIGIPLIKADKV-QAQGFGANVYKVAVDIGIQ 141
QY 172 ANHPDLKGVIGVYDYNVGRTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVYK 231
DB 142 ASHPDL--NVVGCASFVAGEAYNTDGNHGHGTHVAGTVAALDNTTG-VLGVAPSVSLAVK 198
QY 232 VLGDGSGSVSTIIAGVDYVQNKDKYGRVNLNLGLSSOSSDQDTSLSQAVNNNAWDAGI 291
DB 199 VLNSGSGTYSGLVSGIEMATTN---GMDVINMSLG---GPSGSTAMKQAYDNAINGV 251
QY 292 VVCAAGNSGP--NTYVGSPPAAASKVITVGVADSDNDNIASFSSRGTADGRLKPEVYAP 349
DB 252 VVVAAGNSGSGNTIGYPAKYDSVIAYGAVDSNSNRKSFSSVG-----AELEVAP 305
QY 350 GVDIIAPRASGTSGTPIINDYTKASGTSMATPHVSGVGAALILQAHPSWTPDKYKTAI E 409
DB 306 GAGVSTYPTST-----YATLNGTSMASPHVAGAAALILSKHPNLSASQVRNRLSS 356
QY 410 TADIVAPKEIADIAYAGRVNRYKA 434
DB 357 TATYLG----SSFYIGKGLINVEAA 377

Search completed: October 31, 2002, 13:32:26
Job time : 17 secs

GenCore version 5.1.3
 Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:27:48 : Search time 17 seconds
 (without alignments)
 3724.875 Million cell updates/sec

Title: US-09-841-553-1
 Perfect score: 3437
 Sequence: 1 MKRGAVVLAALVGLAGT.....YASYGVADYQLKAVVYG 659

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : PIR-71.1
 1: PIR1:
 2: PIR2:
 3: PIR3:
 4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	669.5	19.5	442	2 A69587	intracellular alkaline
2	637	18.5	444	2 B83891	intracellular alkaline
3	597.5	16.2	379	1 SUBSCL	subtilisin (EC 3.4
4	547	15.9	382	2 I39780	subtilisin (EC 3.4
5	523.5	15.5	382	1 SUBSN	subtilisin (EC 3.4
6	523.5	15.5	1398	2 T28159	pyrolysin (EC 3.4
7	524.5	15.3	274	1 SUBSD	subtilisin (EC 3.4
8	522.5	15.2	806	2 A41341	microbial serine p
9	519.5	15.1	275	2 JC1085	alkaline protease
10	515.5	15.0	384	2 JC4802	alkaline protease
11	513.5	14.9	381	2 JH0778	subtilisin (EC 3.4
12	512.5	14.9	381	1 SUBSI	subtilisin (EC 3.4
13	510.5	14.9	380	2 A49778	high-alkaline seri
14	510	14.8	1433	1 A36734	subtilisin (EC 3.4
15	509.5	14.8	381	1 SUBSS	subtilisin (EC 3.4
16	509.5	14.8	381	2 JQ1487	subtilisin (EC 3.4
17	497.5	14.5	272	2 C84324	subtilisin (EC 3.4
18	497	14.5	757	2 C84120	subtilisin (EC 3.4
19	495	14.4	378	2 A39781	subtilisin-type pr
20	491.5	14.3	378	2 A39773	high-alkaline seri
21	488.5	14.2	279	1 SUMITV	thermitase (EC 3.4
22	482	14.0	401	2 I39974	serine proteinase
23	477.5	13.9	799	2 G83753	subtilisin-type pr
24	473.5	13.8	535	2 B82358	alkaline serine pr
25	473.5	13.8	715	2 JC4908	alkaline serine pr
26	472	13.7	513	1 A35742	aqualysin (EC 3.4
27	470	13.7	361	2 A48373	high-alkaline seri
28	470	13.7	361	2 G83756	subtilisin-type al
29	464.5	13.5	534	1 JS0173	alkaline proteinase

ALIGNMENTS

RESULT 1

A69587

Intracellular alkaline serine proteinase aprx - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: A69587

R:Kunst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

A.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

Nature 390, 249-256, 1997

A:Authors: Foulger, D.R.; Hilbert, A.; Hilbert, K.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningsstein, G.; Krugh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

V. M.; Ogawa, K.; Ogiwara, S.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portate

Rieger, M.; Rivolta, C.; Roche, E.; Roche, M.; Sadale, J.; Sato, T.; Scanl

A:Authors: Schleicher, S.; Schreiber, R.; Scoffone, P.; Sekiguchi, J.; Sekowaka, A.; So

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpatre, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Emswiler, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69587; MUID: 98044033

A:Accession: A69587

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-442 <KUN>

A:Cross-references: GB:299113; ERLAL009126; MID:92634090; PIDN:CAB13610.1; PID:el1833

A:Experimental source: strain

C:Genetics:

A:Gene: aprX

C:Superfamily: subtilisin homology

F:146-398/domain: subtilisin homology <SBT>

Query Match

Best Local Similarity 19.5% Score 669.5; DB 2; Length 442;

Matches 151; Conservative 51; Mismatches 103; Indels 23; Gaps 9;

QY 127 FIEDKVKVDDATSVSGIGDGVNSLQSGGVVAVYDTGIDANHPDLKGVIGWYD 185

DB 116 YLNRVKALLDTATEASH-FAVVRNGOTLTGKGVTVAVYDTGI-YPHDLEGRIGPAD 173

QY 187 AVNGRSTPYDQGGITVAGVAGTGSVNS-QYIGVAPGAKLVGVKVGADGSGSVSTII 245

DB 174 MVQKTPYDNGHGTTCAGI-ASSGASSSGQYRGPAPEANLGVKVKNGKSGTLDII 233

QY 246 AGVDVYVO-NKDK--YGRVIVNLSLGS---QSSDGTGDSLSQAVNANDAGIYCVVAGN 299

DB 234 EGVEWCIQVNEEDPEIDTINSLGLDALRYDHEQEDPLRYAVEASAGIYCVVAGN 293

QY 300 SGNPTVTVGSPAASKVTVTVVNSN-----DNIASSSSRGPTADGRKLAPEVAGVD 352

DB 294 SGPDSTIASPGSEKVTYTGALDDNTASSDDDTVASFSRGRPTVYVYKPKDILAPGVN 353

Wed Nov 6 14:29:31 2002

A:Molecule type: DNA
 A:Residues: 1-379 <JAC>
 A:Cross-references: GB:X03341; NID:9487721; PIDN:CA856500.1; PID:95921206
 A:Experimental source: strain NCI86816
 J:Smith, E.L.; DeLange, R.J.; Evans, W.H.; Landon, M.; Markland, F.S.
 J. Biol. Chem. 243, 2184-2191, 1968
 A:Title: Subtilisin Carlsberg, the complete sequence: comparison with subtilisin B
 A:Reference number: A00968; MUID: 68234702
 A:Accession: A00968
 A:Molecule type: protein
 A:Residues: 106-206, 'S', 208-232, 'A', 234-261, 'N', 263-264, 'S', 266-315, 'N', 317-379 <SMI>
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
 not necessary for normal sporulation.
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:1-29/Domain: signal sequence; status predicted <SIG>
 F:30-103/Domain: propeptide; status predicted <APT>
 F:106-379/Domain: subtilisin Carlsberg; status experimental <NPT>
 F:128-339/Domain: subtilisin homology <SBT>
 F:137,169,325/Active site: Asp, His, Ser; status predicted
 Query Match 16.28; Score 557.5; DB 1; Length 379;
 Best Local Similarity 37.98; Pred. No. 3.8e-23;
 Matches 146; Conservative 56; Mismatches 134; Indels 49; Gaps 13;
 QY 57 NQEVDTVMGSGYCDRDAVK--VLRMGAGVYKIIIPAVAVKIKAROLLIIAGMIDT 114
 DB 35 NVEKDYIVGFKS-GVKTASVKKVQIKESGKVDKQFRIINAAKAKLDKEKKEVKNPDV 93
 QY 115 GFEGNTRVSGIKFTQEDYKVO--DDAT---SVSQTGADTVNNSLGYDQSGVVVAIVDTGID 171
 DB 94 AY-----VEEDRYHALAATVPYGIPLKADKY-QAOGFKGANVAVLDGTG 141
 QY 172 ANHPDLKGVKGTWDAVNGRS--PYDDQGHGTHVAGTGVNSQYIGVAPGAKLVGVK 231
 DB 142 ASHPLD--NVVGGASFVAGVANTDNGEGTHVAGTVAALDNTG-VLGVAPSVSLVAVK 198
 QY 232 VLGDGSGSVSTIAGVDVYVQKDKYIRVINSLSGSSOSSDSDTSLSSQVNNWADAGI 291
 DB 199 VLNSSGSYSGIVSGIEKATN---GNVNNLSG---GPSGSTMKQAVDNRVARGV 251
 QY 292 VVCVAAGNSGP--NYYVGVSPAAASKVITGVAVDSNDNIASFSSRGPTADGRKPEVAP 349
 DB 252 VYVAAAGNSGSGNTNTICIPAK--DSVIANGVAVDSNRSASFSSVG-----AELEVNA 305
 QY 350 GVDIIAPRASGTSMTGPIINDYITARSCTSMATPHYSVGVALILOAHPSWTPDKVKLTALIE 409
 DB 306 GAGVSTYPTST-----YPLNGTSMASPHVAGAAALILSKHPNLSASQVNRLLS 356
 QY 410 TADIVAPKEIADIAVAGANVYKA 434
 DB 357 TATYLG-----SSFYTGKGLINVEAA 377
 RESULT 4
 I33780
 subtilisin (EC 3.4.21.62) Sendai precursor - Bacillus sp.
 C:Species: Bacillus sp.
 C:Date: 19-Jul-1996 #sequence, revision 19-Jul-1996 #text_change 22-Jun-1999
 C:Accession: I33780
 R:Yamagata, Y.; Isshiki, K.; Ito, Shima, E.
 Enzyme Microb. Technol. 17, 643-663, 1995
 A:Title: Subtilisin Sendai from alkalophilic Bacillus sp.: molecular and enzymatic pr
 A:Reference number: I33780; MUID: 95329264
 A:Accession: I33780
 A:Status: preliminary; transcribed from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-382 <RES>
 A:Cross-references: GB:D29688; PIDN:RAA06157.1; PID:9595964
 C:Genetics:
 A:Gene: aprQ
 A:Start codon: TTG
 C:Superfamily: subtilisin; subtilisin homology

353 IIAIPRASGT-----SNGTPIINDYITARSCTSMATPHYSVGVALILOAHPSWTPDKVKLTAL 407
 DB 354 IISLRSPNSYIDKLQKSRVSGQFTMGSYSMTATPICAGIAALILQNPDLTPDEVKELL 413
 QY 408 IETADIVAPKEIADIAVAGANVYKA 435
 DB 414 KNGTD--KWKDEPNLYGAGAVNAENSV 439
 RESULT 2
 B83891
 Intracellular alkaline serine proteinase aprX [imported] - Bacillus halodurans (strain C
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence, revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: B83891
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11036132
 A:Accession: B83891
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-444 <SRO>
 A:Cross-references: GB:AP001513; GB:BA000004; NID:G10174345; PIDN:BA805649.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: aprX
 Query Match 18.58; Score 637; DB 2; Length 444;
 Best Local Similarity 39.28; Pred. No. 2.3e-27;
 Matches 146; Conservative 68; Mismatches 122; Indels 36; Gaps 10;
 QY 85 QVKSYSKIIPAVAVKIKARDL--LLTAGMIDTGYEFTNTRVSGIKFTQEDYKVOVDATSV 142
 DB 82 QLRHSFKIKCSVSNWTPESLOEMLVCKDKRIY-----LNREVALDITAVES 130
 QY 143 SQIQGADTVNNSLGYDQSGVVVAIVDTGIDANIPDLKGVIGWDAVNGRSTPYDDQGHGT 202
 DB 131 AQ-AFEVIRNGETLTGKDYITAVDTGI-YPHEDLEGRIKAFDYVNRQEPYDDNGHGT 188
 QY 203 HVAGTVAGTG-SYNSQYIGVAPGAKLVGVKLGADGSGSVSTIAGVDVYVQKDKY-- 258
 DB 189 HCAGDARGAGASSDGOYRGPAPEANVIGVRLNKGQMGSLGSLMGVENCQIENEIPDD 248
 QY 259 GTRVINSLSGSS--QSSDGTDSLQAVNNWADAGIVVCVAGNSGPNYTYVGSFPAASK 315
 DB 249 PIHISMSLGQALPENEQEDPHVRYVEAMNAGITVCVAGNSGPDAGTIAACPVSSEK 308
 QY 316 VITVGAVD-----SNDNIASFSSRGPTADGRKPEVAVPGVDIIAPRASGT---SM 363
 DB 309 VITVGALDDRTDREDDOAVFSSRGPTIYGRKPKDILAPGVNIVLSRSPNSFYDKTQK 368
 QY 364 GTFINDYTKASCTSMATPHYSVGVALILOAHPSWTPDKVKLTALIEADIVAPKEIADIA 423
 DB 369 GSRVSHYTMNSGTSMATPVCAVWALMLQHEPLTPDEVKTRLMESTDRWADND--PVV 426
 QY 424 YGAGRVNYKAI 435
 DB 427 YGAGYISAEGAI 438
 RESULT 3
 SUBSCL
 subtilisin (EC 3.4.21.62) Carlsberg precursor - Bacillus licheniformis
 C:Species: Bacillus licheniformis
 C:Date: 30-Jun-1988 #sequence, revision 30-Jun-1988 #text_change 21-Jul-2000
 C:Accession: A24111; A00968
 R:Jacobs, M.; Eliasson, M.; Uhlen, M.; Flock, J.I.
 Nucleic Acids Res. 13, 8913-8926, 1985
 A:Title: Cloning, sequencing and expression of subtilisin Carlsberg from Bacillus lichen
 A:Reference number: A24111; MUID:86093688
 A:Accession: A24111

C:Keywords: hydrolase; serine proteinase
F:136-342/Domain: subtilisin homology <SBT>

Query Match 15.98; Score 547; DB 2; Length 382;
Best Local Similarity 34.58; Pred. No. 1.4e-22;
Matches 134; Conservative 73; Mismatches 145; Indels 14; Gaps 17;

QY 1 MKRLGAVL---ALVVLGLAGTALAAPVKPVRRNNAOQNYGLLTPGLFKKVPQNNH-57
DB 1 MKRLFKVYASALLLSLTSATSVSAE-----EQKQYLI--GFENQLQVTEPV 48
QY 58 QEVD-----TVINFGSYGDRDRAYKVLRLMGAQVKYKIIIPAVAVKIKARDLLIAGKD 113
DB 49 ESSDKQSQMSLFAEYDESIEMELL-----YEPEDIPVVSVELSPVKDL----- 95
QY 114 YFGFNTVRVSGIKIQDYKQVDDATS---VSGIADTVNWSLGYDGSQVVAIVDTGI 70
DB 96 -----ENDPSITVIEDIVITNOVTPMGITRVOAPTAN--TRGITGTGRVAVLDTGT 148
QY 171 DANHPDLKGVGYDQVAVNGRSTPYDDQGHGTHVAGTGVSNQSYIGVAPGAKLVG 230
DB 149 -STHPDL--NIRGVSEVPQEPSYQDGNHGHGTHVAGTIAALNN-SIGVGVAPNAELAY 204
QY 231 KVLGADGSGYSTIAGVQWQKDKYIRVNLISGSSQSSDGTDSLQAVNNAWDAG 290
DB 205 KVLGANGSGSVSSIAOGLQWTAQN---NIHVNLSLGSVP---GQSTLELAVNOATNAG 257
QY 291 IVCVVAAGSGPNTYVGVPAASKVITGVGAVSDNINAFSSRSRPTADGRLKPEVAPG 350
DB 258 VLWVAATNGSG--TVSPARYANALVAGTADONNNRASFSQYGTGLN-----IVAPG 309
QY 351 VDIAPRASGTSGTPIINDYTKASGTSMATPHVSGVAGLILQAHPSMTDKVKTALIE 410
DB 310 VGI-----GSTYPGNRYASLSGTSMATPHVAGVAALVKQKNPSMTQLRHLIST 360
QY 411 ADIVAPKEIADIAGAGRVNVYKAIK 436
DB 361 ATSLGNSN----QFGSLVNAEAAVR 382
RESULT 5
SUBSN
subtilisin (EC 3.4.21.62) BPN' precursor - Bacillus amyloliquefaciens
N:Alternate names: subtilisin Novo
C:Species: Bacillus amyloliquefaciens
C:Date: 24-Apr-1984 #sequence_revision 28-Aug-1985 #text_change 21-Jul-2000
C:Accession: B25415; A93495; T44584; A92033; A00970
R:Vasantha, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Nagle, J.; Filipina, D.
J. Bacteriol. 159, 811-819, 1984
A:Title: Genes for alkaline protease and neutral protease from Bacillus amyloliquefaciens
A:Reference number: A25415; MUID:85006739
A:Accession: B25415
A:Molecule type: DNA
A:Residues: 1-382 <VAS>
A:Cross-references: GB:K02496; NID:gl42525; PIDN:AAB05345.1; PID:gl42526
R:Wells, J.A.; Ferrara, E.; Henner, D.J.; Estell, D.A.; Chen, E.Y.
Nucleic Acids Res. 11, 7811-7925, 1983
A:Title: Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in
A:Reference number: A93495; MUID:84069812
A:Accession: A93495
A:Molecule type: DNA
A:Residues: 1-382 <VEL>
A:Accession: T44584
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: W.8-382 <NE2>
A:Cross-references: EMBL:X00165; NID:g39337; PIDN:CAA24990.1; PID:g775500-
J. Biol. Chem. 242, 5198-5211, 1967
A:Title: Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complete
A:Reference number: A92033; MUID:68086682
A:Accession: A92033

A:Molecule type: protein
A:Residues: 108-162, 'PN', 155-167, 'D', 169-194, 'SA', 197-204, 'DA', 207-264, 'ST', 267-357, 'R'; Kaut, J.
In The Enzymes, 3rd ed., vol.3. Boyer, P.D., ed., pp.547-560, Academic Press, New York
A:Title: Subtilisin: X-ray structure.
A:Reference number: A94443
C:Contents: annotation: X-ray crystallography, 2.5 angstroms; active site
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
not necessary for normal sporulation.
C:Genetics:
A:Start codon: GTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-32/Domain: signal sequence; status predicted <SIG>
F:33-107/Domain: activation peptide; status predicted <APT>
F:108-382/Product: subtilisin #PN; status experimental <MPT>
F:130-342/Domain: subtilisin homology <SBT>
F:139-171, 328/Active site: Asp, His, Ser; status experimental

Query Match 15.58; Score 533.5; DB 1; Length 382;
Best Local Similarity 38.58; Pred. No. 7.7e-22;
Matches 143; Conservative 52; Mismatches 113; Indels 63; Gaps 14;
QY 78 VLRLMGAQVKYKIIIPAV-----AVKIKARDLLIAGHIDTGYGNTVRVSGTKFIQZ 130
DB 59 VISEKGGKVKQKPYDAAASATNEKAVKELKD-----PSVAYVEE 100
QY 131 DYKQVDDAT---SVSQIGQVTVNWSLGYDGSQVVAIVDTGIDANHPDLKGVIGWYDA 187
DB 101 DHVAVAYAQSPVGVQIKAPAL-HSQGYTGSVKVAVVDSGIDSSHPLD--KVAGASM 157
QY 188 VNGRSTPY-DDQGHGTHVAGTGVSNQSYIGVAPGAKLVGVKLGADGSGSVSTIFA 246
DB 158 VPSETNPFQDNHSHGTHVAC--VAALNN-SIGVLGVAPASLYAVKVLGADGSGQVSWIN 216
QY 247 GVDVWVONKDKYIRVNLISGSSQSSDGTDSLQAVNNAWDAGIYVYVAGNSGP--NT 304
DB 217 GIENAIAN---NMDVINSLG---GPSGSAKAAKAVKAVAGSVVYVAAAGNECTSS 269
QY 305 YTVGSPAAASKVITGVAVDSNDINAFSSRSRPTADGRLKPEVAPGVDTIAPRASGT 364
DB 270 STVGYDCKTPSVIAVGVAVDSSNORASFSSVGPELD-----VMAPGVSI-----QS 314
QY 365 TPINDYTKASGTSMATPHVSGVAGLILQAHPSMTDKVKTALIEADIVAPKEIAD-IA 423
DB 315 TLPCKNYGATNGTSMASPHVAGAAALILSKHPMTWTVQVRSLENTT-----TKLQDSFY 369
QY 424 YGAGRVNVYKA 434
DB 370 YGKGLINVAQA 380

RESULT 6

T38159
Pyrolysin (EC 3.4.4.1) - Pyrococcus furiosus
C:Species: Pyrococcus furiosus
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 24-Oct-2000
R:Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Plattewouw, C.; Slazen, R.J.; Vo
J. Biol. Chem. 271, 20436-20431, 1996
A:Title: Isolation and characterization of the hyperthermostable serine protease, pyr
A:Reference number: 220481; MUID:96355370
A:Accession: T38159
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1398 <VOO>
A:Cross-references: EMBL:U55835; NID:gl556462; PID:gl556463; PIDN:AAB09761.1
A:Experimental source: DSM3638
C:Genetics:
A:Gene: pls
C:Keywords: hydrolase; serine proteinase
Query Match 15.58; Score 533.5; DB 2; Length 1398;

Db 206 YSTPTNTYAT-LNGTSMASHVAGAAALILSKHPNLSASQVRNRLSSTATYLG----SS:363

Qy 422 IAGAGRVNYKA 434

Db 261 FYGKGLINVEAA 273

RESULT 10

JC4802 alkaline proteinase (EC 3.4.21.-) precursor - Thermoactinomyces sp. (strain E79)

C:Species: Thermoactinomyces sp.

A:Variety: strain E79

C:Date: 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change 09-Jun-2002

C:Accession: JC4802

R:Lee, J.K.; Kim, Y.O.; Kim, H.K.; Park, Y.S.; Oh, T.K.

Biosci. Biotechnol. Biochem. 60, 840-846, 1996

A:Title: Purification and characterization of a thermostable alkaline protease from Ther

A:Reference number: JC4802; MUID:96261070

A:Accession: JC4802

A:Molecule type: DNA

A:Residues: 1-384 <LEE>

A:Cross-references: GB:U31759; NID:g1389689; PIDN:AAB36499.1; PID:g1f83629

A:Experimental source: strain E79

C:Comment: This protein is thermostable.

C:Function: extracellular alkaline serine proteinase [validated. MUID:96261070]

C:Superfamily: subtilisin; subtilisin homology

C:Keywords: hydrolase; serine proteinase

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-106/Domain: propeptide #status predicted <PRO>

F:107-383/Product: alkaline proteinase #status experimental <MAT>

F:134-344/Domain: subtilisin homology <SBT>

F:143-176/Active site: Asp, His, Ser #status predicted

Query Match 15.0% Score 515.5; DB 2; Length 384;

Best Local Similarity 32.9% Pred. No. 7.3e-21;

Matches 147; Conservative 77; Mismatches 150; Indels 73; Gaps 13;

Qy 1 MKRLGAVLALVYGLLA--GTALAAKPKVVRNNAVOQKNYGLTLPCLPKVQVRMNHQ 58

Db 1 MKRFLSVATLLVLLAVFTPEAA--SPASDDXVPCELLVFKDGISAQ-----50

Qy 59 EYDTVMFGSGYDRDRAVYKRLMGAQVYKSIPIPAVAVIKARDLLIAGMIDIGYRG 118

Db 51 --STQSIHAQYG--AKSIEKSYLGFV-----VKPDGS----VEKNIEK-YKA 49

Qy 119 NTRVSGIKRPIQEDYKVOY----DDATG---VSQIGADTVNLSLGYDGSVYVAIYDTQZ 170

Db 91 NPWV---EYVEPNHYVHLMTPTNDLTSRWGPKVKQAPQAWD-VTRSSSTVIAIVDTQW 145

Qy 171 DANHPDLCKVIGWDAVNGSTPYDQGHGTHVAGIVAGTGSVNSOYIGVAPGAKLVG 210

Db 147 QTHNPDLQKQVGGDFVNDNSNPQDNGHGTHCAIAAANTNGTGIAGHAPASINPV 206

Qy 231 KVLGADSGSVSTIIAGYDVMYQNKYIRVINSLSGSSSSDGTDLGQAVNNAWDAC 290

Db 207 RVLNNSGSGTMAAVANGIAYAAQN---GADVLSLSG---GTSGSSALASAVQQAWSG 259

Qy 291 IVCVVAAGSGNPVTYVGSPPAAKSVITVYGAVDNSNDNTASPSRCPTADGRKPEVYVAG 350

Db 260 AVVVAAGNS--SSSTPNPAYVQAATAVASTDNDLSLFSYNSGWD-----VAAEG 311

Qy 351 VDIIAPRASGTSMGTPINDYTKASGTSMATPHVSCVGGALLIQAHPSPWPKVKVKTALIE 410

Db 312 SNIVS-----TYLNSVYASLSGTSMATPHVAGLAALL--ASQGRNSQIRAAIENT 360

Qy 411 ADIVAPKEADIAYGAGRVNYKA 437

Db 361 ADKISG---TGTFQGRINAYKAVNY 384

RESULT 11

JH0778

subtilisin (EC 3.4.21.62) NAT precursor - Bacillus subtilis (strain natto NC2-1)

N:Alternate names: natto proteinase; nattokinase; subtilisin Bsp

C:Species: Bacillus subtilis

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000

C:Accession: JH0778; JS0601; JSC617; JC2036

R:Nakamura, T.; Yamagata, Y.; Tsuchishima, E.

Biosci. Biotechnol. Biochem. 56, 1869-1871, 1992

A:Title: Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus subtilis (

A:Reference number: JH0778; MUID:93113095

A:Accession: JH0778

A:Molecule type: DNA

A:Residues: 1-381 <NAK>

A:Cross-references: GB:D25319; NID:g435439; PIDN:BAA04989.1; PID:g435440

R:Sumi, H.

Kagaku To Seibutsu 29, 119-123, 1991

A:Title: Natto kinase and fibrinolysis.

A:Reference number: JS0601

A:Accession: JS0601

A:Molecule type: protein

A:Residues: 107-381 <SUM>

R:Sumi, H.; Nakajima, N.

Nippon Nogeikagaku Kaishi 65, 125-1127, 1991

A:Title: Studies on fibrinolysis enzymes in fermentation food.

A:Reference number: JS0517

A:Accession: JS0517

A:Molecule type: protein

A:Residues: 107-381 <FUD>

C:Genetics:

A:Gene: aprN

A:Start codon: GTG

C:Superfamily: subtilisin; subtilisin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-106/Domain: activation peptide #status predicted <PRO>

F:107-381/Product: subtilisin NAT #status experimental <MAT>

F:139-341/Domain: subtilisin homology <SBT>

F:118,139,170,327/Active site: Asp, Ser, His, Ser #status predicted

Query Match 14.3% Score 513.5; DB 2; Length 381;

Best Local Similarity 38.1% Pred. No. 9.2e-21;

Matches 139; Conservative 52; Mismatches 124; Indels 49; Gaps 13;

Qy 78 VLRLMGAQVYKSIPIPAVAVIKARDLLIAGMIDTGYFGNTRVSGIKFTQEDYKQVD 137

Db 58 VISEKGGKVKQFKYVNAATLDEKAVKEL-----KKDPSVAVVEEDH-LAHE 105

Qy 138 DATSV-----SQIGADTVNNSLYDGSVYVAIVDTGIDANHPDLAKVIGWDAVNGRST 193

Db 106 YAOSVPTGISQIKAPAL-HSGSYTGSNVKVAVIDSGIDSSHDDL--NVRGASVFPSETN 162

Qy 194 PYDD-QGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVYKVLGADGSGSVSTIIAGYDMVY 252

Db 163 PYDGSSTHGHVAGTIALNN-SIGLVGAPASLVAVKVLSTGSGQYSWIINGIEWAI 221

Qy 253 QNRDKYIRVINSLSGSSSGDGLSQAANNADAGIVCVCAAGNSP--NTYTVGSP 310

Db 222 SN-----NMVDVNNLSG-----PTGSLTKVTYDRAVSGIVAAAAGNSGSGSTVGVYP 274

Qy 311 AAASKVITVCAVDNSNDNTASPSRCPTADGRKPEVYVAPGVVDIIAPRASGTSMGTINDY 370

Db 275 AKYPTSTIAGAVNNSNQRAFP--GVGSELD-----VNAQVSIQTSLPGCT-----319

Qy 371 YTRASGTSMATPHVSGVGLIQAHPSPWPKVKVKTALIEADIVAPKEADIAYGAGRVN 430

Db 320 YGAYNGTSMATPHVAGAAALILHPTWTNAQVDRLESTATYLG-----NSFYIGKGLIN 375

QY 431 VYKA 434

Db 376 VQAA 379

RESULT 12

SUBSI

subtilisin (EC 3.4.21.62) B precursor - Bacillus subtilis
 N:Alternate names: alkaline proteinase; bacillopeptidase E; extracellular alkaline serin
 C:Species: Bacillus subtilis
 C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-34-2000
 C:Accession: A00972; A26116; I39970; I39778; S68012; H69586
 R:Stahl, M.L.; Ferrari, E.
 J. Bacteriol. 158, 411-418, 1984
 A:Title: Replacement of the Bacillus subtilis subtilisin structural gene with an in vitro
 A:Reference number: A00972; NUID:84212198
 A:Accession: A00972
 A:Molecule type: DNA
 A:Residues: 1-381 <STA>
 A:Cross-references: GB:K01988; NID:g143519; PIDN:AAA22742.1; PID:g143520
 A:Experimental source: strain 1168
 R:Mong, S.L.; Price, C.W.; Goldfarb, D.S.; Doi, R.H.
 Proc. Natl. Acad. Sci. U.S.A. 81, 1184-1188, 1984
 A:Title: The subtilisin E gene of Bacillus subtilis is transcribed from a sigma37 promoter
 A:Reference number: A26116; NUID:84144862
 A:Accession: A26116
 A:Molecule type: DNA
 A:Residues: 1-155 <NON>
 A:Cross-references: GB:K01443; NID:g143665; PIDN:AAA22814.1; PID:g143666
 R:Ikemura, H.; Takegi, H.; Inouye, M.
 J. Biol. Chem. 262, 7859-7864, 1987
 A:Title: Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli
 A:Reference number: I39978; NUID:87222417
 A:Accession: I39978
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-156 <YKE>
 A:Cross-references: GB:M15639; NID:g143521; PIDN:AAA22744.1; PID:g143523
 R:Hegner, D.J.; Ferrari, E.; Perego, M.; Hoch, J.A.
 J. Bacteriol. 170, 296-300, 1988
 A:Title: Location of the targets of the hpr-97, sacU32(Hy), and sacO36(Hy) mutations in the hpr operon of Bacillus subtilis
 A:Reference number: I39778; NUID:88086885
 A:Accession: I39778
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-8 <HEN>
 A:Cross-references: GB:M19125; NID:g143527; PIDN:AAA22245.1; PID:g143528
 R:Park, S.
 J. Bacteriol. 171, 2657-2665, 1989
 A:Title: Bacillus subtilis subtilisin gene (aprE) is expressed from a sigma-A (sigma-43) promoter
 A:Reference number: I39779; NUID:89213955
 A:Accession: I39779
 A:Molecule type: DNA
 A:Residues: 1-13 <PAR>
 A:Cross-references: GB:M31060; NID:g143529; PIDN:AAA22246.1; PID:g143530
 A:Experimental source: strain M166, substrain PV79
 R:Kanai, M.; Hoeoeg, J.O.; Kaiser, R.; Shafiq, J.; Razzaki, T.; Zaidi, Z.H.; Joernvall, J.
 FEBS Lett. 374, 363-366, 1995
 A:Title: Isolation, characterization and structure of subtilisin from a thermotolerant Bacillus subtilis strain
 A:Reference number: S68012; NUID:96069945
 A:Accession: S68012
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 113-323 <NAM>
 R:Kunat, F.; Ogawara, N.
 C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chidambaram, P.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabrat, C.; Ferrari, E.; Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galled
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.;
 A:Authors: Leuber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maudel
 y, M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelid

Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadalo, Y.; Sato, T.; Scani
 A:Authors: Schleich, S.; Schloeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; So
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchlya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Zumbach, E.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; NUID:98044033
 A:Accession: H69586
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-26,197,281 <KUM>
 A:Cross-references: GB:299108; GB:AL009136; NID:g2633260; PIDN:CAB12870.1; PID:g26333
 A:Experimental source: strain 168
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
 not necessary for normal sporulation.
 C:Genetics:
 A:Gene: aprE
 A:Map position: 690-771
 A:Start codon: GTG
 C:Function:

A:Description: catalyzes the hydrolysis of peptide bonds
 A:Note: this enzyme has broad specificity and will hydrolyze peptide amides; it prefe
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; protein digestion; serine proteinase
 F:1-23/Domain: signal sequence (status predicted <Sig>
 F:24-106/Domain: activation peptide (status predicted <Act>
 F:107-381/Product: subtilisin E (status predicted <APT>
 F:129-341/Domain: subtilisin homology <SDT>
 F:138,170,327/Active site: As5, His, Ser (status predicted

Query Match 14.9% Score 512.5; DB 1; Length 381;
 Best Local Similarity 37.9%; Pred. No. 1e-20;
 Matches 138; Conservative: 53; Mismatches 124; Indels 49; Gaps 13;

QY 78 VLRLMGAGVYKVKITPRVAVKAKADLLIAGNIDTGFGRVSGIKFIOEDYKQVVD 137
 Db 58 VISEKGVKQKQKRYVAAATLDEKAVKE-----KKDPSVAVYEDH-IAHE 105
 QY 138 DNTSY----SOIGADIV+NSLQ+DQSGVVAIVDTGIDANHPDLKRGKVGWYAVNGRST 193
 Db 106 YAOVYPYGIQIKAPAL+HSGG+TGSNKVAVIDSGIDSSHPDL--NVRGASVYSEIN 162
 QY 194 PYDD-OGHTRGVAGVAGT+SVISQYIGVAPGAKLVGVKVLGADSGSVSTIAGVDWVY 252
 Db 163 PYDQSSNGTHVAGTIA+LW+SIGVLGVSPLASLYAVKYLDSGSGOYSHINGIENAI 221
 QY 253 ONKDKYGIKRVINILSGSSVSEGTDSLSQAVNMDAGIVYCVAGNSGP--NTYTVGSP 310
 Db 222 SN---NMVYNNISLG--++GTGTAUKTVDKAVSSGIWAAAGAGSGSGSTSTVGYP 274
 QY 311 AAASKVITVGVDSNDNIATP+SRGPTADGRKPEVAVPQVDIAPRASGTSMGTPINDY 370
 Db 275 AKYPTIANGAVNSNQRA+ESSAGSELD-----VMAPGVSIQSTLPQGT----- 319
 QY 371 YTKASGTSMATPHVSGVCAELGAPHPSTDPKVKTLLETADIYAPKEIADIYAGGRVW 430
 Db 320 YGAYNGTSMATPHVAGAA+KSKKHPTWNNQVRDLSESTATYLG-----NSFTYTGAGLIN 375
 QY 431 VYKA 434
 Db 376 VQAA 379

RESULT 13

A49778

N:Alternate names: subtilisin homology, high-alkaline
 C:Species: Bacillus alcalophilus
 C>Date: 12-Mar-1994 #sequence_revision 24-Feb-1995 #text_change 20-Jun-2000
 C:Accession: A49778; JCI244
 R:van der Laan, J.C.; Gerritse, G.; Mulleners, L.J.S.M.; van der Hoeek, R.A.C.; Quax,
 Appl. Environ. Microbiol. 57, 901-909, 1991

A:Title: Cloning, characterization, and multiple chromosomal integration of a bacillus

A:Reference number: A49778; MUID:91282483

A:Accession: A49778

A:Molecule type: DNA

A:Residues: 1-380 <N>

A:Cross-references: GB:M5086; NID:g142456; PIDN:AAA22212.1; PID:g142457

A:Experimental source: strain PB92, ATCC 31408

A:Note: amino acid of mature protein confirmed by peptide sequencing

R:Kobayashi, T.; Kobayashi, M.; Yamamoto, M.; Nakamura, S.; Aono, R.; Horike

Biosci. Biotechnol. Biochem. 56, 1455-1460, 1992

A:Title: Molecular cloning, nucleotide sequence, and expression of the structural gene

A:Reference number: JCI1244; MUID:93043753

A:Accession: JCI1244

A:Molecule type: DNA

A:Residues: 1-195, 'S', 197-380 <T>

A:Cross-references: GB:D13157; NID:g216231; PIDN:BA02443.1; PID:g216232

A:Experimental source: Bacillus alcalophilus Vedder, ATCC 21522 (Bacillus sp. 221)

A:Superfamily: subtilisin, subtilisin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-27/Domain: signal sequence; status predicted <SIG>

F:112-380/Product: alkaline serine proteinase; status predicted <PRO>

F:134-340/Domain: subtilisin homology <SBT>

F:143-173,326/Active site: Asp, His, Ser #status predicted

Query Match 14.9%; Score 510.5; DB 2; Length 380;

Best Local Similarity 32.9%; Pred. No. 1.3e-20;

Matches 146; Conservative 75; Mismatches 148; Indels 75; Gaps 17;

2 KRLGAVLVLV-----GLAGTAAAPKPVYNNVAVQKNGLLTGLPKKQVRN 55

Db 3 KPLGKINASTALLISVAFSSSIASAAEKYILGFNQE-----AVSEFEQVE 53

QY 56 WQNEVDVTINFGSGDRDRAVKVRLMGAOVKYSKTIIPAVAVKAKRDLLIAGMIDG 315

Db 54 ANDEVAIV-----SEEEVEI-----ELLHEFETIPVLSVPEVDVALE--LD-- 96

QY 116 YFGRVRSKIFQEDKVOVDAT---SVQIGADVTNLSLGDGSGVVAIVDTGIDA 172

Db 97 -----PAISYIEDEAVTTAAQSPVNGISRVQAPAAHNR-CLTGSGVKVAVLDGI-S 147

QY 173 NHPDLKGVIGNYDAVNGRSTPYDDGHTGHTVAGTGSVNSQYIGVAPGKLVGVV 232

Db 148 THPDL--NIRGASFPVGPSTODGNGHTGHTVAGTIAALNN-SIGVLGVAPNEXYAVY 204

QY 233 LGADGSGVSTIIAGVQVYQVNDKXGIRVNLISGSSQSDGDTLSQVNNARDAGIV 252

Db 205 LGASGSGVSSIAQGLEWAGN---GMEVANLSLGSFSPS---ATLEQVNSATSRGL 257

QY 293 VCVAGNSGPNYTVGSPARAKVITVGAVDSDNDIASFSRGTADGRKPEVVAPEVD 332

Db 258 VVAASGNSGAG--SISIPARYANNAVATDNNNRASFSQYAGLD-----IVAGSVN 309

QY 353 IIAPRASGTSMTPTINDYTKASGTSNATPHVSGVGLIIQNPSTPKYKTLLEAD 412

Db 310 VOSTYPGST-----YASLNGTSNATPHVAGAAVVKQNPSPNSVOIRNHLKNTAT 360

QY 413 IVAPKEIADYAGVGRVYKAIK 436

Db 361 SLGSTNL-----YGSGLVNAEATR 380

RESULT 14

A36734

bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000

C:Accession: A36734; A35131; A35750; B35750; S08223; JN0335; B3456; JUC084

R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.

J. Bacteriol. 172, 5520-5521, 1990

A:Reference number: A36734; MUID:90368623

A:Contents: erratum

A:Accession: A36734

A:Molecule type: DNA

A:Residues: 1-1433 <SLO>

A:Cross-references: GB:M29035; NID:g143307; PIDN:AAA2679.1; PID:g143308

R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.

J. Bacteriol. 172, 1470-1477, 1990

A:Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and clo

A:Reference number: A35131; MUID:90170864

A:Accession: A35131

A:Molecule type: DNA

A:Residues: 1-365, 'S', 367-882, 'EIMP', 893, 'Q', 895-896 <SL2>

A:Cross-references: GB:M29035

A:Note: the authors translated the codon GAA for residue 545 as Leu

R:Wu, X.C.; Nathoo, S.; Pang, S.H.; Carne, T.; Wong, S.L.

J. Biol. Chem. 265, 6845-6850, 1990

A:Title: Cloning, genetic organization, and characterization of a structural gene enc

A:Reference number: A35750; MUID:90216713

A:Accession: A35750

A:Molecule type: DNA

A:Residues: 876-935, 'CG', <WU2>

A:Cross-references: GB:J05400; NID:g142607; PIDN:AAA83363.1; PID:g1119197

A:Note: this sequence has been corrected

R:Masuda, E.S.; Anaguchi, R.; Satoh, T.; Takeuchi, M.; Kobayashi, Y.

Nucleic Acids Res. 19, 657, 1991

A:Title: Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus subtilis.

A:Reference number: S08223; MUID:90174995

A:Accession: S08223

A:Molecule type: DNA

A:Residues: 1410-1433 <NAS>

A:Cross-references: EMBL:X17344; NID:g40165; PIDN:CAA35224.1; PID:g809661

R:Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.

Biochim. Biophys. Acta. 56, 1166-1168, 1992

A:Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectri

A:Reference number: JN0335; MUID:93005071

A:Accession: JN0335

A:Molecule type: protein

A:Residues: 195-218, 'A', <KAT>

A:Note: source of this material was Bacillus subtilis (natto)

R:Beall, B.; Lowe, M.; Lutkenhaus, J.

J. Bacteriol. 170, 4855-4864, 1988

A:Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia co

A:Reference number: I39849; MUID:89008108

A:Accession: I39849

A:Status: preliminary; translated from GB/EHBL/DBB

A:Molecule type: DNA

A:Residues: 1-211 <RES>

A:Cross-references: GB:M226370; NID:g142938; PIDN:AAA22458.1; PID:g551705

R:Kunst, F.; Ogawara, N.; Hozier, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Saitoh, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Puma, S.; Galizzi, A.; Gal

lach, J.; Harwood, C.R.; Henaut, S.; Hilbert, H.; Holsappel, S.; Hosono, S.; Kullu, M

Koetter, P.; Koningsstein, G.; Kragh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y.; M.; Ogawa, K.; Ogilva, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yaeumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Yamamoto, H.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:9804033

A:Accession: B69596

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1433 <KUN>

A:Cross-references: GB:29911; GB:299112; GB:AL009126; NID:g2633907; PIDN:CAB13404.1;

A:Experimental source: strain 53

C:Genetics:

A:Gene: bpr; bpf
 A:Map position: 135 (degrees)
 C:Superfamily: bacillopeptidase F; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:1-30/Domain: signal sequence status predicted <SIG>
 F:31-194/Domain: propeptide status predicted <PRO>
 F:195-1433/Product: bacillopeptidase F status experimental <MAT>
 F:218-466/Domain: subtilisin homology <SBT>
 F:227,274,452/Active site: Asp, His, Ser #status predicted

Query Match 14.8%; Score 510; DB 1; Length 1433;
 Best Local Similarity 26.4%; Pred. No. 8.3e-20;
 Matches 210; Conservative 104; Mismatches 302; Indels 178; Gaps 32;

QY 3 RLGAVALVGLVGLL-----AGTALAAPKVPVVR-----NAVQKNGYGLTGLPKKVCOR 53
 DB 8 RLISVLSVTVVYSSLPFGAAGASSKVTSPVKKELQSAESIONKISSLSKKS-PKKKEK 66
 QY 54 MNMNOEVDTVIMFGSGYGDRAVKVL-----RLMGAQVKYSYKIIPAVAVKIKAA----- 102
 DB 67 TTF-----LIKQDLANPEKAARAAKAKSKLSAAKTEYQKRAVYSSLSKYTADQSO 120
 QY 103 RDLILLI-----AGMIDTGY-----PKN 119
 DB 121 QDLVLYLTKQKGNADQIHSTYVNGTAVRASKEVMEKVQPEVEKVLPLNEKRLQFKS 180
 QY 120 TVNSGKIQEDYKQVQDDATSVSIGADTVNSLGYDGGVYVAIVDTGIDANHPDLKG 179
 DB 181 SFFPNMKAQRAIKATDGVEMVYDQIDAPKAW-ALGYDGTGTVASIDTGVEMNHPALKE 239
 QY 180 KVIQ-----HYDAVNGRSTPYDDGHTGTHVAGIVAGTGVNSQYIGVAPGA 205
 DB 240 KYRGYNPNPNPEPENENWYDAVAGEASPYDLAGHTHTVGTWYGEPTQNGVAPGA 299
 QY 226 KLUGVKVUGADGGSGSVSTIAGVQVWVQNKDKY-----IRVNLGSGSQSDGQDS 278
 DB 300 KVIAYKAFSEDG-GTQADILEAGWYLPAPKAQGNPHENAPDVVNSWG-----GGSGLDE 355
 QY 279 LSOAYNNAWADAG-IVVCVAAGNS-----GPNYTVGSPAAASKVITVCAVSDNDNIASF 331
 DB 356 WYRDVNAWNAADIFFEFSAGNTDLFIPGPG--SIANPANYEPESATCATDINKKLAQ 413
 QY 332 SSRGPTADRLKPEVYAPGVDDIAPRASGTSKGTPIINDYTKASGTSNATPHVSGVAL 361
 DB 414 SLGGSPYDEIKPEISAPGVNI---RSS-----VPGQTYEDGHDGTSNAGPHVSAVAL 465
 QY 392 LQAHPSWTPPKVKYKTALTIETADIVAPKEADI---AYGAGRVNYYKAIKYDDYAKULTPS 468
 DB 466 KQANASLVDEMEDILTSTAELTDSTFPDSPNNGYCHGLVNAFDV-----SA 514
 QY 449 VAD-KGSATHTFDVSG-----ATFVATLYWDTSDDLDLYDPNGNEVDYSYANY 500
 DB 515 VTDGLGKAGQVSEGDQPPYQHEKVTBAYEGSLPLTLPAD--NVSVTSYKLSY 571
 QY 501 GPEKGVYNNPTA-----GTWTVKVVYKGAANYQVDVSD--GSLSSQSGGNGNPN 549
 DB 572 KLDCQGETEITAKRISDHLKGTQAEIPDKTKLSYKWNHDFGCHVSSDNYDVTYK 631
 QY 550 PNPNTPTPTDTQ-----FTGSVNDY-WDTSDFTTNVNSG-----ATKITGDLTFTSY 598
 DB 632 PSITAGYKQDEPAPGCVWASGNTNNWCVPTSGPTAASGEKYYVGLTGLGNYA----- 686
 QY 599 NOLDLYLDP-----NGNLVDRSTSSNS-----YENVEYANPAPGTWFL-VYAYSTY 645
 DB 687 NSANNLVMPPIKAPQSGSLQFKSWHNLEDDEFDYGVFLPEGEKKNWQAGVYNGKTS 746
 QY 646 GADYQKAVVYVG 659
 DB 747 SNTDEIDLSAYKG 760

RESULT 15

SUBSS

subtilisin (EC 3.4.21.62) amylosaccharitic precursor - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 24-Apr-1984 sequence revision 24-Feb-1995 #text_change 16-Jun-2000
 C:Accession: A11448; A00971; S68013
 R:Yoshimoto, T.; Oyama, H.; Honda, T.; Tone, H.; Takeshita, T.; Kamiyama, T.; Tsuru, J. Biochem. 103, 1060-1065, 1968
 A:Title: Cloning and expression of subtilisin amylosaccharitic gene.
 A:Reference number: A11448; MUID:89008194
 A:Accession: A11448
 A:Molecule type: DNA
 A:Residues: 1-381 <YOS>
 A:Cross-references: GB:000264; NID:9216328; PIDN:BA00186.1; PID:9912425
 R:Experimental source: var. amylosacchariticus
 R:Kurihara, M.; Markland, F.S.; Smith, E.L.
 J. Biol. Chem. 247, 5619-5631, 1972
 A:Title: Subtilisin amylosacchariticus. III. Isolation and sequence of the chymotrypt
 A:Reference number: A00971; MUID:72266688
 A:Accession: A00971
 A:Molecule type: protein
 A:Residues: 107-112-114-148-152-155-157-164-170-173-174-178-181-200-205-210-212-219-2
 R:Experimental source: var. amylosacchariticus
 R:Kamali, M.; Hoeber, J.O.; Katsner, R.; Shafiqat, J.; Razzaqi, T.; Zaidi, Z.H.; Joernva
 FEBS Lett. 374, 363-366, 1995
 A:Title: Isolation, characterization and structure of subtilisin from a thermostable
 A:Reference number: S68012; MUID:96069945
 A:Accession: S68013
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 107-235-27, 237-345-293-381 <KAW>
 C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
 not necessary for normal sporulation.
 C:Genetics:
 A:Start codon: GTC
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:107-381/Product: subtilisin #status predicted <MAT>
 F:129-341/Domain: subtilisin homology <SBT>
 F:138,170,327/Active site: Asp, His, Ser #status predicted

Query Match 14.68; Score 509.5; DB 1; Length 381;
 Best Local Similarity 37.34; Pred. No. 1.5e-20;
 Matches 137; Conservative 54; Mismatches 124; Indels 49; Gaps 13;

QY 78 VLRLMGAQVKYSYKIIPAVA KIKARDLLIAGMIDTGYEGMTRVSGIKFIQEDYKQVOD 137
 DB 58 VISEKGGVKQKQPKYVNAAGAYLDEKAVKL-----KKQPSVAYVEEDH-IAHE 105
 QY 138 DATSV-----SOIGADTVNLSLGYDGGVYVAIVDTGIDANHPDLKQKVGTHDAYNGRST 193
 DB 106 YAQSVFYGIQIKAPAL-HSQYITGSNKKVAVIDSGIDSSHDL--NVRGGASFPVSETN 162
 QY 194 PYDD-QGHGTHVAGIVAGTGVNSQYIGVAPGAKLVGKVLGADGGSGSVSTIAGVQVYV 252
 DB 163 PYQDGGSHGTHVAGTTAALNT-SIGLVGPSASLYAVKVLDTSGSGQYSWIIINGENAI 221
 QY 253 QNKQKYGTRVNLISLASSQSLGTDLSQAVNNANDAGIVVCAAGNSCP--NTYVCGSP 310
 DB 222 SN-----NWDVNNLSG---GPSGTALKTVDVDAVSSGIVVAAAAGNEGSSSTVGY 274
 QY 311 AAASKVITVCAVSDNDNIATSSRGPTADGRKPEVYAPGVDDIAPRASGTSNMGTPINDY 370
 DB 275 AKYPSITAVGAVNNSNORAFSSAGSELD-----VMAPGVSTQSTLPQGT----- 319
 QY 371 YTKASGTSNATPHVSGVGLIILQAHPSWTPDKVKYKTALTIETADIVAPKEADIADIVAGRVN 430
 DB 320 YGAYNNTSMATPHVAGAAAILLSKHPTWTNAQVRDRLESTATYLG-----NSFYKGLIN 375
 QY 431 YKKA 434
 DB 376 VQAA 379

Wed Nov 6 14:29:31 2002

us-09-841-553-1.rpr

Page 10

Search completed: October 31, 2002, 13:31:55
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:25:28 ; Search time 10 seconds

(without alignments)
2551.618 Million cell updates/sec

Title: US-09-841-553-1

Perfect score: 3437

Sequence: 1 MRLGAVLVLLVGLLAGT.....YAYSTYGVADYQLKAVVYVS 639.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	557.5	16.2	379	1	SUBT_BACLI
2	533.5	15.5	382	1	SUBT_BACAM
3	533.5	15.5	1398	1	PLS_PYRFU
4	524.5	15.3	274	1	SUBD_BACLI
5	522.5	15.2	806	1	SUBV_BACSU
6	513.5	14.9	381	1	SUBN_BACNA
7	512.5	14.9	381	1	SUBT_BACSU
8	510.5	14.9	380	1	ELVA_BACAO
9	510.5	14.9	380	1	ELVA_BACCS
10	510.5	14.8	1433	1	SUBF_BACSU
11	509.5	14.8	381	1	SUBT_BACSA
12	509.5	14.5	275	1	SUBT_BACST
13	499.5	14.5	275	1	SUBT_BACPU
14	491.5	14.3	378	1	ELVA_BACSP
15	488.5	14.2	269	1	PRPM_BACSP
16	488.5	14.2	279	1	THET_BACPU
17	487.5	14.2	269	1	SUBS_BACLE
18	485.5	14.1	269	1	SUBB_BACLE
19	482	14.0	401	1	THES_BACSP
20	476	13.8	422	1	TKSU_PYRKO
21	472	13.7	513	1	AQLI_THEAQ
22	470	13.7	361	1	ELVA_BACHD
23	464.5	13.5	534	1	PROA_VIBAL
24	447	13.0	408	1	SEPR_THESR
25	437.5	12.7	645	1	SURE_BACSU
26	412.5	12.0	467	1	ISP6_SCHPO
27	399.5	11.5	326	1	ISP_PAEPO
28	399	11.6	339	1	ISPI_BACSU
29	391	11.4	395	1	BPRX_BACNO
30	350	11.3	580	1	PROA_XANCP
31	386.5	11.2	478	1	YSP3_YEAST
32	386.5	11.2	535	1	SPN1_MAGCR
33	384	11.2	894	1	WPRA_BACSU

34	378	11.0	385	1	CUDP_METAN
35	375.5	10.9	635	1	PRTB_YEAST
36	375	10.9	403	1	ALP_GSPAC
37	372.5	10.8	1902	1	P3P_LACLC
38	372	10.8	426	1	SUBT_BAC99
39	372	10.8	533	1	HLX_HALLI7
40	370.5	10.8	1902	1	PIP_LACLC
41	370	10.8	382	1	PRTR_TRIAL
42	370	10.8	1902	1	P2P_LACPA
43	369	10.7	384	1	PRTK_TRIAL
44	367.5	10.7	321	1	ISP_BACCS
45	365.5	10.6	533	1	PEPC_ASPNG

ALIGNMENTS

RESULT 1

SUBT_BACLI

ID SUBT_BACLI STANDARD: PRT: 379 AA

AC P00780:

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Subtilisin Carlsberg precursor (EC 3.4.21.62).

GN APR.

OS Bacillus licheniformis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1402;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIB 6816;

RX MEDLINE=86093688; PubMed=5001653;

RA Jacobs M., Eliasson M., Uhlen M., Flock J.-I.;

RT "Cloning, sequencing and expression of subtilisin from

RT Bacillus licheniformis."

RL Nucleic Acids Res. 13:8913-8926(1985).

RN [2]

RP SEQUENCE OF 106-379.

RX MEDLINE=68234702; PubMed=4957581;

RA Smith E.L., Delange R.J., Evans W.H., Landon M., Markland F.S.;

RT "Subtilisin Carlsberg. V. The complete sequence; comparison with

RT subtilisin BPN'; evolutionary relationships."

RL J. Biol. Chem. 243:2184-2191(1968).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF VARIANT WITH SELENOCYST-325.

RX MEDLINE=93291170; PubMed=8512925;

RA Syed R., Wu Z.-P., Hogle J.M., Hilvert D.;

RT "Crystal structure of an encapsubtilisin at 2.0-A resolution."

RL Biochemistry 32:6157-6164(1993).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 106-379.

RX MEDLINE=98087517; PubMed=9425066;

RA Stoll V.S., Eger B.T., Eger R.C., Martichonok V., Jones J.B.,

RA Pai E.F.;

RT "Differences in binding sites of enantiomers of 1-acetamidoboronic

RT acid based protease inhibitors: crystal structures of gamma-

RT chymotrypsin and subtilisin Carlsberg complexes."

RL Biochemistry 37:451-462(1998).

CC [1] FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,

CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.

CC [2] CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity

CC for peptide bonds, and a preference for a large uncharged residue

CC in pI. Hydrolyses peptides and

CC [3] SUBCELLULAR LOCATION: Secreted.

CC [4] BIOTECHNOLOGY: used as a detergent protease. Sold under the name

CC Alcalase by Novozymes.

CC [5] MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF

CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY

CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN

CC IS NOT NECESSARY FOR NORMAL SPOULATION.

CC [6] SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

FT TURN 267 268
 FT STRAND 274 276
 FT TURN 275 277
 FT TURN 279 280
 FT TURN 282 287
 FT TURN 289 290
 FT TURN 293 293
 FT TURN 295 296
 FT TURN 301 302
 FT TURN 305 308
 FT TURN 312 316
 FT TURN 317 319
 FT STRAND 320 324
 FT STRAND 327 344
 FT TURN 346 347
 FT TURN 350 359
 FT TURN 360 360
 FT STRAND 362 362
 FT STRAND 367 370
 FT TURN 371 372
 FT STRAND 374 374
 FT TURN 377 381
 FT HELIX 382 AA; 39181 MW; ED987DAPA37B8335 CRC64;
 SQ SEQUENCE 382 AA; 39181 MW; ED987DAPA37B8335 CRC64;
 Query Match 15.5%; Score 533.5; DB 1; Length 382;
 Best Local Similarity 38.5%; Pred. No. 3.1e-23;
 Matches 143; Conservative 52; Mismatches 113; Indels 63; Gaps 14;
 QY 78 VLRMGAVKYSYKIIIPAV-----AVKIKARDELLIAGMIDRFGNTRVSGIKFICE 130
 DB 59 VISERGGKVKQFYKVDYDAAGATLNERAKRELKKD-----PSVAYVEE 100
 QY 131 DYKVVQDDAT---SVSGIGADTVNSLGYDGVVVAIVDGTIDANIPDLKGVKIGHYDA 187
 DB 101 DRYAHAYAGSPYGVYQKAPAL-HSGGTGTSNVAIVYVDSIGDSSHPDL--KYAGGASM 157
 QY 188 VNGRSTPY-DOQGHOTHYAGVAGTSVNSQYIGVAPGAKLVGVKLGADSGSGSVSTIIA 246
 DB 158 VPSETNPFODNNSHTHAGTVAAALNN-SIGVLGVAPASLYAYKVLGADSGSQYSWIIN 216
 QY 247 GVDWYQNKDKYGLRVINLSLGSQSSDGTDSLQAVNNAVDAGIYCYVAGNSGP--NT 304
 DB 217 GIEAIAI-----NMDVINMSLC---GPSGSALKAAVDKAVASGVVVAAGNECTSGSS 359
 QY 305 YTVGSPAASKYITVGVADSNNTASFSRGPTADGRUKPEVYAPGVDDIIPASGTSMG 354
 DB 270 STVGPCKYPSVIAVGVDSNQRASFSSVGPELD-----VMAPGVSI-----QS 314
 QY 365 TPINDYTRASGTSNATPHSVGVCALTLQAHPSMTPOKVKTLTETADIVAPKEIAD-IA 423
 DB 315 TLPCKYKAYNGTSMASPHVAGAAALILSKHPNWTNQVRSLENTT-----TKLGDSFY 369
 QY 424 YGAGRVNYYKA 434
 DB 370 YGKGLINYQAA 380
 RESULT 3
 ID PLS PYRFU STANDARD; PRT; 1398 AA.
 AC PZ186; 2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Pyrolysin precursor (EC 3.4.21.-).
 GN PLS OR PF0287.
 OS Pyrococcus furiosus.
 OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=2261;
 RN SEQUENCE FROM N.A., SEQUENCE OF 150-184, AND CHARACTERIZATION.
 RC STRAIN=VCI / DNS 3638;
 RX MEDLINE=96355370; PUBMED=8702780;

RA Voorhorst W.G.B., Eggen R.A.L., Geerling A.C.M., Platteuw C.,
 RA Siezen R.J., de Vos W.M.;
 RA "Isolation and characterization of the hyperthermostable serine
 RA protease, pyrolysin, and its gene from the hyperthermophilic archaeon
 RA Pyrococcus furiosus.";
 RL J. Biol. Chem. 271:20426-20431(1996).
 RN [2]
 RP CHARACTERIZATION, AND 3D-STRUCTURE MODELING.
 RX MEDLINE=21079021; PubMed=11210516;
 RA de Vos W.M., Voorhorst W.G.B., Dijkgraaf M., Kluskens L.D.,
 RA Van der Oost J., Siezen R.J.;
 RA "Purification, characterization, and molecular modeling of pyrolysin
 RA and other extracellular thermostable serine proteases from
 RA hyperthermophilic microorganisms.";
 RL Meth. Enzymol. 330:383-393(2001).
 CC -1- FUNCTION: Has endopeptidase activity toward caseins, casein
 CC fragments including alpha-SI-casein and synthetic peptides.
 CC -1- SUBCELLULAR LOCATION: Cell-envelope associated.
 CC -1- PTM: LWM pyrolysin seems to be produced by autoproteolytic
 CC activation of HMW pyrolysin.
 CC -1- PTM: Glycosylated.
 CC -1- MISCELLANEOUS: Thermostable; high activity at 95 degrees Celsius.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC This SWISS-PROT entry is CopRight. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC ENBL: U55835; AAB03761.1;
 CC HSSP: 045670; 1DBI.
 DR HEROPS; S08 100;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 4.
 DR PRINTS: PRO0723; SUBTILISIP.
 DR PROSITE: PS00136; SUBTILASE_1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 ?
 FT PROPEP 1 ?
 FT CHAIN 149 1398 PYROLYSIN
 FT ACT_SITE 150 1398 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 365 365 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 152 152 N-LINKED (PROBABLE).
 FT CARBOHYD 222 222 N-LINKED (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (POTENTIAL).
 FT CARBOHYD 240 240 N-LINKED (POTENTIAL).
 FT CARBOHYD 257 257 N-LINKED (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (POTENTIAL).
 FT CARBOHYD 298 298 N-LINKED (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (POTENTIAL).
 FT CARBOHYD 551 551 N-LINKED (POTENTIAL).
 FT CARBOHYD 663 663 N-LINKED (POTENTIAL).
 FT CARBOHYD 739 739 N-LINKED (POTENTIAL).
 FT CARBOHYD 792 792 N-LINKED (POTENTIAL).
 FT CARBOHYD 893 893 N-LINKED (POTENTIAL).
 FT CARBOHYD 908 908 N-LINKED (POTENTIAL).
 FT CARBOHYD 917 917 N-LINKED (POTENTIAL).
 FT CARBOHYD 929 929 N-LINKED (POTENTIAL).
 FT CARBOHYD 1048 1048 N-LINKED (POTENTIAL).
 FT CARBOHYD 1056 1056 N-LINKED (POTENTIAL).
 FT CARBOHYD 1084 1084 N-LINKED (POTENTIAL).
 FT CARBOHYD 1117 1117 N-LINKED (POTENTIAL).
 FT CARBOHYD 1133 1133 N-LINKED (POTENTIAL).
 FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
 FT CARBOHYD 1148 1148 N-LINKED (POTENTIAL).

FT CARBOHYD 1208 1208 N-LINKED (POTENTIAL).
FT CARBOHYD 1233 1233 N-LINKED (POTENTIAL).
FT CARBOHYD 1237 1237 N-LINKED (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (POTENTIAL).
SQ SEQUENCE 1398 AA; 154495 MW; 69E49C4A23C3F824 CRC64;
Query Match 15 59; Score 533.5; DB 1; Length 1398;
Best Local Similarity 22.69; Pred. No. 1.6e-22;
Matches 239; Conservative 104; Mismatches 248; Indels 467; Gaps 37;
Qy 2 KRIGAVVLAIVLGLAGTALAAPVKPVYRNAAVQ--KNYGLLTGFLKFKVQRMN-- 57
Db 4 KGLTVLFIAMLSVVPVHFVSAGTPPVSSSENSTSLPNOQVW----KEVSAALNAI 59
Qy 58 ---QEVDTVMGSGDRDAVKVRLKGAQVYSKIIIPAVAKI---KARDLLLAGM 111
Db 60 MKQCPNNVLIITKKGLEAKTELEKLAGIEIDENRYNMLLYKIKPKVKELNYISL 119
Qy 112 IDTGFCNTRVSGIKFIQEDKVKQVDDATSVSGIGADTYNSL-----YDGSQVYVAI 165
Db 120 EKAWLAREVKLSP-PIVEKDVKTR-EPSLEPKMYNSTWVINALQFIQBFYDGSQVVA 177
Qy 166 VDTGIDANR-----POLKGVIGWYD-----AVNG----- 190
Db 178 LDTGVDPNPFSLITPDGRRKLIENKFTDEGFVDFSFYSKVVNGLIINTTQVASGL 237
Qy 191 ----- 190
Db 238 TLNESTGLMEYVYVTVSVNTIGNITSANGIVHFLGLPERYFDLNFQDQEDFPVLLV 297
Qy 191 RST----- 193
Db 298 NSTGNGDYATYVDTDLDTDFDPLQCYNWTYDVAVFYSYGGPLNYVLAIDPNSYAV 357
Qy 194 -PYDDGCHTHVAGIVAGTGSV-----SOYI-----GVAP 223
Db 358 FGWDGCHGCHTHVAGIVAGTGSVSDNDADWLSYSGEMEFSLYGDYNTVTTDTCGVAP 417
Qy 224 GAKLVGVKVLGADSGSVSTIAGVDMVQNKQYGRVNLISL-GSSQSSDGTDSLSQA 282
Db 418 GAOIMAIRVLRSRGRSMDIIEGTAA---THADVISLSLGNAPYLDGTDPSVA 473
Qy 283 YNANDR-GIVCYAAGNSGNTYTVGSPAAASKVITVGAVD----- 323
Db 474 VDELTKEYGVVFIAGNEGPGIINVGSPGATKAITGAAAVPINVGVSQALGVPYV 533
Qy 324 -----SDNIIASFSGRTADGRKPEVAVPGVDIIAPRASGTSNGTPIINDYVTKA 374
Db 534 YGFYFPAYTNVRIAFSSRGPRIDGEIKPNVAVPGYI---SSLPNWIGGADF--- 386
Qy 375 SGTSMATPHVSGVAGALIOA-HPS---WTPDKVKTALETA-----DIVAPKEIADIAYC 425
Db 587 SGTSMATPHVSGVAVALLISGPKPEGIYINPDIIKKVLESQATWLEGDPYTCQKYTELDQ 486
Qy 426 AGRVNV-----YKAI-----KYDDYAKL----- 442
Db 647 HGLVNVTKSWEILKALINGTTLPIVDHNAKSDYSDFAEYLGVDVIRGLYARNISPDIVEM 706
Qy 444 -----TF-----TGSV----- 449
Db 707 IKVIGOTEYRTEIVATEPMKPFVSGSVILENTEFVLRVKYDVEGLEPGLYVGRIID 766
Qy 450 -----ADKGSATHT-FDVSGATEVT-----ATLY 472
Db 767 DPTTPVIEILNTIVIEPKTEPBNVTLTWYDINGPENVTHTFTVPEGVVDLYAMTT 836
Qy 473 WDTGSSDIDLXYDPNG-----NEVDYSYATAYGPEKGVYNTACTVTVKVSYKGAAN 527
Db 827 WDTG-----LIRPDGFVFPYQLDYLPA-----VSNPNGNRELVTGTFNAPC 871
Qy 528 YQVDVVSQSLSCSGGNNPNPNPNPTTDTQTGTGSVN-DYWDTSDTFTMNVN----- 582
Db 872 YE-----SGFLVRINGVEITPS-----VWVINTYLDNTNTEFSIEFNITNI 912

Qy 583 -----SG-TNITGDLTP-----DTSYNDDLDLYL 605
Db 913 YAPINATLIPIGLGTYNASVSGDEFFKIEVPEGTAEIAIRIGNSVPNSDLDLYL 972
Qy 606 YDPNGNLVDRSTSSNYSYHVEANPAGTWTFLVAYS 643
Db 973 YDSKGNLVALDGNPTAEETVVEYKPKGVYSIVVHGYS 1010

RESULT 4

SUBD: BACLI
ID SUBD: BACLI STANDARD; PRT: 274 AA.
AC P00781.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Subtilisin DY (EC 3.4.21.52).
GN APR.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE.
RC STRAIN-DY;
RX MEDLINE=84110064; PubMed=6420308;
RA Nedkov P., Oberthur W., Braunitzer G.;
RT "Primary structure of subtilisin DY";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1537-1540(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN-DY;
RX MEDLINE=99041548; PubMed=9826175;
RA Eschenburg S., Genov N., Peters K., Fittkau S., Stoeva S.,
RA Wilson K.S., Betzel C.;
RT "Crystal structure of subtilisin DY, a random mutant of subtilisin
Carlabers";
RL Eur. J. Biochem. 257:309-318(1998).
CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds, and a preference for a large uncharged residue
in P1. Hydrolyses peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
SPORULATION, AND MANY MUTATIONS WHICH BLOCK SPORULATION AT EARLY
STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPORULATION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.
DR PIR; A00969; SUBSD.
DR PDB; 1DH6; 18-NOV-98.
DR MEROPS; S08.001.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASASP; 1.
DR PROSITE; PS00137; SUBTILAS_HIS; 1.
DR PROSITE; PS00138; SUBTILAS_SER; 1.
KW Hydrolyase; 32 32 CHARGE RELAY SYSTEM.
FT ACT_SITE; 32 32 CHARGE RELAY SYSTEM.
FT ACT_SITE; 63 63 CHARGE RELAY SYSTEM.
FT ACT_SITE; 220 220 CHARGE RELAY SYSTEM.
SQ SEQUENCE 274 AA; 27435 MW; 0154696E22F46333 CRC64;

Query Match 15 59; Score 524.5; DB 1; Length 274;

Best Local Similarity 42.28; Pred. No. 6.6e-23;

Matches 132; Conservative 38; Mismatches 94; Indels 49; Gaps 10;

Qy 124 GKKEOEDYKVOVDDATSVSGIADTVNNSGVDGSGVVAIVDTGTIDANHPDLAKGVIG 183

Db 7 GIPLIKAD-KVQAQ-----GYKANVKVGIIDTGLAASHTDL--RVVG 46


```
QY 184 MYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQYIGVAPCAKLVGVKVLGADSGSVST 243
DB 47 GASFVSGESYNTDNGHGTHVAGTVAALDNTTG-VLGVAPVNSLYAIAKVLNSSGSGTVA 105
QY 244 IAGVDDVWQKQYIGYRVLNLSLSSOSSDGTSLSQVNNANDAGIVVYVAGNSGSP 302
DB 106 IVSGIEWATQ-----GLDVNLSG---GPGSGTALQAQVKAAYGIVVYVAGNSGSS 159
QY 303 -NTYTVGSPAAASKVIYVAGVSDNNDIASFSRGTADGRLKPEVAVPGVDIAPRSGM 361
DB 159 GSNTIGYVPAKYDSVIAVAGVSDNNDIASFSRGTADGRLKPEVAVPGVDIAPRSGM 212
QY 362 SNGTINDYTKASGTSMATPHVSGALILQAHPSTPKVKATLIEADIVAPKEIAD 421
DB 213 -----YVSLNGTSMATPHVSGALILQAHPSTPKVKATLIEADIVAPKEIAD 259
QY 422 IAYGAGRVNYYKA 434
DB 260 FYVKGGLNVEA 272

RESULT 5
SUBV_BACSU STANDARD; PRT; 806 AA.
AC F29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Minor extracellular protease VPR precursor (EC 3.4.21.1).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
RN RP MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer J.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT Mb region from 325 degrees to 333 degrees."
RL Mol. Microbiol. 10:371-384(1993).
CC -1- FUNCTION: NOT REQUIRED FOR GROWTH OR SPOULATION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in its way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/en/home
CC or send an email to license@isb-sib.ch).
```

```
DR Subtilist; BG10591; vpr.
DR InterPro; IPR003137; PA.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILASIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Ymogen; Signal; Complete proteome.
FT SIGNAL 28
FT PROPEP 29
FT CHAIN 160
FT ACT_SITE 161
FT ACT_SITE 189
FT ACT_SITE 233
FT ACT_SITE 534
FT ACT_SITE 534
SQ SEQUENCE 806 AA; 85603 MW; F984E3BF0865DDDD CRC64;

Query Match 15.24; Score 522.5; DB 1; Length 806;
Best Local Similarity 24.74; Pred. No. 3.3e-22;
Matches 190; Conservative 102; Mismatches 208; Indels 269; Gaps 30;

QY 72 RORAVKVLRLMGAQVYKY-KI-PAVAVKIKAROLLIAIMID-----TGYFCNTRV 122
DB 90 KNAIKAVK--NGKVNREYETSGFSMKLPANEIPKLLAVKDKVAVPNVYKTDNMDK 147
QY 123 SGKFTQEDYKQVDDATSVSGADTVNSLGYDGSVYVAVYDTCIDANHPDLK---G 179
DB 148 KDVTSIEDAVSPOMDD--SAPYIGANDWD-LGYTGKGIKVAIIDGVYNHDPDKKNGF 204
QY 180 KVTGWDAVNG-----RSTPY70-----QGRGTHVAGIVAGTGSVNSQYIGVAPCAKLVG 230
DB 205 QYKG-YDFVNDYDPKETPTGDPGEATDGHGTHVAGTVAANGTIK---GVAPDATLLAY 259
QY 231 KVLGADSGSVSTIAGVDRVYQNKDKYGIYRVLNLSLSSOSSDGTSLSQVNNANDAG 290
DB 260 RVLGPGSGGTENYJAGVERAQD-----GADVNNLSLNSUNPDW-ATSTALDWAHSEG 314
QY 291 IVYVAGNSGPNYTYGSAFAASKVITYGA-----VDSNDN----- 327
DB 315 VVAVTNSGSPNGTWTGATGATGIRALISVGAQLPLNEYAVTFGYSYSAKVMGINKEDDV 374
QY 322 ----- 327
DB 375 KALNKEVELVEAGIGEMAFESKDLTGKVVAVVKGSGTAFVOKADNAKAGAGIMVYNN 434
QY 328 -----IASFSRGP 336
DB 435 LSGETEANYFGNSVPTINKLEDEKLVSAKAGETKTFKLTYSKALGOVADFSRGP 494
QY 337 TAD-GRLEKPEVYVAPGVDTIAPRASGTSMTGTPINDY-YTKASGTSMTATPHVSGVGLILQA 394
DB 495 VMDTNMIKPDISAQVNT-----STIPTHDPHPYGYSGKSGTSMASPHIAGAVAVIKQA 550
QY 395 HPSTWTDKVKATLIEADIVAPKEIADY-----GAGRVNYYKAIKYDDYAKLTFTGSV 449
DB 551 KPAKWSVEQIKKAIMNTA-----LKDSDGCEVPHNAOGASARINNAIKADSL----- 599
QY 450 ADKGSATHTFEDVSGATFVYATHTDGTSSDIDLILYDPNGNEVDYSTAYYGEKGVYNN 509
DB 600 -----VSPGSGVSTGKENGNETKNETFTIENOSIRKSYTLEYSPNGSGI-- 646
QY 510 PIAGTWTYVAVSYK--GAANYC-VYVSDGSLSSGSGGNPNPNPNPTTDTTFTGSV- 567
DB 647 STSGTSRVVPAHOTGKATAKVVV-----NTKTKAGTYEGTVI 685
QY 568 -----TSDTDTFTMNVNSGATKITGDLTFDT-----SYNDLDLY 604
DB 686 VREGGKTVAKVPTLLIVKEPVY---PRVTSVSYSEGS--VQCTYQIETPLPAGAEELAF 740
QY 605 LYDPNGNLVDRS-----TSSN-TGVEYAN-----PAGTWTFLVYA 641
DB 741 VYDSNLDPAQAGAYIKNQDKT-YFDMDGTINGGTKLPA-GEYLLAYA 788
```



```

CC EMBL: K01988; AA22742.1;
DR PIR: A00972; SUBS1.
DR PDB: 1SCJ; 13-JAN-99.
DR MEROPS: S08.001; -.
DR Subtilisin; BG10190; aprE.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00139; SUBTILASE_SER; 1.
DR Hydrolyase; Spolulation; Serine protease; Zymogen; Signal;
KW 3D-structure; Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN E.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM.
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM.
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM.
SQ SEQUENCE 381 AA; 39495 MW; 52269FAC22E4824F CRC64;

Query Match 14.9%; Score 512.5; DB 1; Length 381;
Best Local Similarity 37.9%; Pred. No. 4.7e-22;
Matches 138; Conservative 53; Mismatches 124; Indels 49; Gaps 13;

QY 78 VLRLMGAQVYKTIIPAVAVKIKAROLLIIAGNIDTGYFGNTRVSGIKTIQEDYKVQVD 137
DB 58 VISEGGRVQKQFYNNAAATLDEKAVKL-----KKDPSVAYVEEDH-IAHE 105
QY 138 DATSV-----SOIGADTVNMSLSDGSGVVAIVDTGIDANHDLKGVKIVGWDAVNGRST 193
DB 105 YACSVPTGICIKAPAL-HSQGTGTSNKKVAVIDSGIDSDHPL--NVRGASFPVSEIN-162
QY 194 PTDD-OCCHTVAGVAGTGVNSQYIGVAPGAKLVGVKVLGADGSGSVSTIAGVDWVY 452
DB 163 PYDQSSGHTVAGTIIAALNN-SIGVLGSPSASLYAKVLDSTGSGOYSMIINGIEMAL 121
QY 253 QNKDYGIRVNLISLGSQSSDGTSLQAVNNADAGIVCVVAGNNSGP--NTYTVGSP 310
DB 222 SN-----NMDVINMSLG---GFTGSTATKTVYDKAVSGIVVAAAGNNGSGSSTVGVYP 274
QY 311 AAASKVITVGVDSNDNISFSSRGPTADGRKPEVYAPQVDIAPRACSTSMKTPINDY 370
DB 275 AKYPSIIAGVAVNSNORASFSSAGSELD-----VAPGVYSTQLPQGT----- 319
QY 371 YTKASGTSNATHSGVGCALILQAPSWTPDKVKTALITADIAPKETAIDYAGGRVN 430
DB 370 YGATNGTSMATPHVAGAAALILSKHTPTWNAQVRDELESTATYLG-----NSFTYCKGLIN 375
QY 431 VYKA 434
DB 376 VQAA 375

RESULT 8
ID ELTA_BACAO STANDARD; PRT; 380 AA.
AC P27693;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus alcalophilus.
OC Bacteria; Firmicutes; Bacillus/clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1445;
RN (1)
RC STRAIN=PB92;
RX van der Laan J.C., Gerritse G., Mulleners L.J.M., van der Hoeft P.A.,
RA Quax W.J.

*Cloning, characterization, and multiple chromosomal integration of a
Bacillus alkaline protease gene.*
Appl. Environ. Microbiol. 57:901-909(1991).
[2]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN=PB92;
RX MEDLINE=92390330; PubMed=1518788;
RA van der Laan J.C., Tepiyakov A.V., Kelders H., Kalk K.H., Misset O.,
Mulleners L.J.M., Dijkstra B.W.;
RT "Crystal structure of the high-alkaline serine protease PB92 from
Bacillus alcalophilus."
RL Protein Eng. 5:405-411(1992).
[3]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RX MEDLINE=93078250; PubMed=144775;
RA Sobek H., Hecht H.-J., Ahle-W., Schomburg D.;
RT "X-ray structure determination and comparison of two crystal forms of
a variant (Asn137) of the alkaline protease from Bacillus
alcalophilus refined at 1.85 Å resolution."
RL J. Mol. Biol. 228:108-117(1992).
[4]
RP STRUCTURE BY NMR OF 112-380.
RC STRAIN=PB92;
RX MEDLINE=97272737; PubMed=911341;
RA Martin J.R., Mulder F.A., Karimi-Mejad Y., van der Zwan J.,
Mariani M., Schipper D., Soelens R.;
RT "The solution structure of serine protease PB92 from Bacillus
alcalophilus presents a rigid fold with a flexible substrate-binding
site."
RL Structure 5:521-532(1997).
CC -!- SUBCELLULAR LOCATION: Secreted
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC ENBL: M65086; AAA22212.1;
DR EMBL: A13738; CAA01128.1;
DR PIR: A49778; A49778.
DR PDB: 1AH2; 15-APR-98.
DR MEROPS: S08.001;
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR Hydrolyase; Serine protease; Zymogen; Signal; 3D-structure.
KW SIGNAL 1 27 POTENTIAL.
FT PROPEP 23 112 ALKALINE PROTEASE.
FT CHAIN 113 380 CHARGE RELAY SYSTEM.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM.
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM.
SQ SEQUENCE 380 AA; 38855 MW; 539EA72771B6682C CRC64;

Query Match 14.9%; Score 510.5; DB 1; Length 380;
Best Local Similarity 37.9%; Pred. No. 6.1e-22;
Matches 145; Conservative 53; Mismatches 148; Indels 75; Gaps 17;

QY 2 KRLGAVVYLVV-----GLIYALAPVPPVNNVAVQKNYGLTGLPKYKVRAN 55
DB 3 KPLGKIVASTALLISVAFSSSTPAAEAEKERYLIGFNEQE-----AVSEFVEQVE 53
QY 56 WNOEVDVTFMGSYGDRQYAVVRLMGAQVYKTIIPAVAVKIKAROLLIIAGNIDTGY 115
DB 54 ANDEVAIL-----SELTEVEI-----ELLHEFTIPVLSVELSPEDVALE--LD-- 96

```

QY 116 YFGNTRVSGIKFQEDYKQVDDAT---SVSGIADTVNLSGDSGVVVAIVDTGIDA 172
 DB 97 -----PAISYIEEDAETTTMAQSPVPGISRVQAPAAHNR-GLTSGGVKVAVLDTGI-S 147
 QY 173 NHPDLCKVIGWYDAYNGRSTPYDDOQGHGTHVAGIVAGTGSVNSQYIGVAPGAKLVGVK 232
 DB 148 THPDL--NIRGASGFVEGPESTQDQNGHGHVAGTIAALNN-SIGVLGVAPNAELVAVKY 204
 QY 233 LGADSGSVSTIAGVDWYVONKRYGIRVINLSGSSQSSDCTDLSQAVNNWADAGIV 292
 DB 205 LGASGSGSVSYIAOGLWAGN---GMHVANLSLGSPPS---ATLEQAVNSATSRGVL 257
 QY 293 VCAVANGSGNTTYVCSPAASKVITYCAVDSNDNTASFSRCPTRADGLKPEVYVAGVD 352
 DB 258 VVAASGNSGAG--SISYPARYANAVGATDONNNRASFSOYGAGLD-----IVAPGVN 309
 QY 353 IIAPRASGTSMTGPIINDYTKASGTSNATPHVSGVGLILQAHPSWTDPKVKTALIEDAT 412
 DB 310 VOSTYPGST-----YASLNGTSMATPHVAGAAALVKQKNPSNRYOIRNHLKNTAT 360
 QY 413 IVAPKEIADIAGAGRVVYKAIK 436
 DB 361 SUGSTNL-----YSGLVNAAEATR 380

RESULT 9
 ID ELYA_BACCS STANDARD; PRT: 380 AA.
 AC P41362;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alkaline protease precursor (EC 3.4.21.-).
 OS Bacillus clausii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=79880;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RX MEDLINE=93043753; PubMed=1368952;
 RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakanura S.,
 Aono R., Horikoshi K.;
 RT "Molecular cloning, nucleotide sequence, and expression of the
 structural gene for alkaline serine protease from alkaliphilic
 Bacillus sp. 221.";
 RL Biosci. Biotechnol. Biochem. 56:1455-1460(1992).
 RN [2]
 RP SEQUENCE OF 112-179.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RA Horikoshi K.;
 RL (In) Horikoshi K. (eds.);
 RL Microorganisms in alkaline environments, pp.187-194, VCH,
 Weinheim (1991).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 SUBTILASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; S48754; AAC60420.1;
 DR EMBL; D13157; BAA02442.1;
 DR EMBL; A26817; CAA01836.1;
 DR EMBL; A22550; CAA01611.1;
 DR HSSP; P29600; IGCI.
 DR MEROPS; S08.103; .

DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; zymogen; signal.
 FT SIGNAL 1 27
 FT PROPEP 28 111
 FT CHAIN 112 380 ALKALINE PROTEASE.
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SO SEQUENCE 380 AA; 38826 MW; 5F73AC68D5B6831 CRC64;

Query Match 14.58; Score 510.5; DB 1; Length 380;
 Best Local Similarity 32.9%; Pred. No. 6.1e-22;
 Matches 146; Conservative 75; Mismatches 148; Indels 75; Gaps 17;

QY 2 KRLGAVLALV-----CLLAGTALAAPVVPVYRNNAVOCKNYGLLTPGLEKKVQRNN 55
 DB 3 KPLGKIVASTALLSVAFSSSIASAAEAKYLLIGFNEOE-----AVSEFVEOYE 53
 QY 56 WNEVDVTVMFGSTGDRRAVKVRLMGAVQVYKIIIPAVAYKIKARDULLIAGIDTG 115
 DB 54 ANDEVAIL-----SEEREVE-----ELLREFETIPVLSVELSPEDYDALE--LD-- 96
 QY 116 YFGNTRVSGIKFQEDYKQVDDAT---SVSGIADTVNLSGDSGVVVAIVDTGIDA 172
 DB 97 -----PAISYIEEDAETTTMAQSPVPGISRVQAPAAHNR-GLTSGGVKVAVLDTGI-S 147
 QY 173 NHPDLCKVIGWYDAYNGRSTPYDDOQGHGTHVAGTGSVNSQYIGVAPGAKLVGVK 232
 DB 148 THPDL--NIRGASGFVEGPESTQDQNGHGHVAGTIAALNN-SIGVLGVAPNAELVAVKY 204
 QY 233 LGADSGSVSTIAGVDWYVONKRYGIRVINLSGSSQSSDCTDLSQAVNNWADAGIV 292
 DB 205 LGASGSGSVSYIAOGLWAGN---GMHVANLSLGSPPS---ATLEQAVNSATSRGVL 257
 QY 293 VCAVANGSGNTTYVCSPAASKVITYCAVDSNDNTASFSRCPTRADGLKPEVYVAGVD 352
 DB 258 VVAASGNSGAG--SISYPARYANAVGATDONNNRASFSOYGAGLD-----IVAPGVN 309
 QY 353 IIAPRASGTSMTGPIINDYTKASGTSNATPHVSGVGLILQAHPSWTDPKVKTALIEDAT 412
 DB 310 VOSTYPGST-----YASLNGTSMATPHVAGAAALVKQKNPSNRYOIRNHLKNTAT 360
 QY 413 IVAPKEIADIAGAGRVVYKAIK 436
 DB 361 SUGSTNL-----YSGLVNAAEATR 380

RESULT 10
 ID SUBF_BACSU STANDARD; PRT: 1433 AA.
 AC P16357;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
 DE (90 kDa serine proteinase).
 CN BPR OR BPF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.; AND SEQUENCE OF 195-222.
 RX MEDLINE=90170864; PubMed=2106512;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
 Pero J.;
 RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
 and cloning of the gene";

RL J. Bacteriol. 172:1470-1477(1990).
 RN [2]
 RN REVISIONS.
 RC STRAIN=168;
 RX STRAIN=90368623; PubMed=2118514;
 RA Slona A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
 RA Pero J.;
 RL J. Bacteriol. 172:5520-5521(1990).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90216713; PubMed=2108961;
 RX Wu X.-C., Nachoc S., Pang A.S.-H., Carne T., Wang S.-L.;
 RA "Cloning, genetic organization, and characterization of a structural
 RT gene encoding bacillopeptidase F from *Bacillus subtilis*.";
 RL J. Biol. Chem. 265:6845-6850(1990).
 RN [4]
 RN SEQUENCE OF 1-211 FROM N.A.
 RP MEDLINE=89008108; PubMed=3139638;
 RX Beall B., Lowe M., Lurkenhaus J.;
 RA "Cloning and characterization of *Bacillus subtilis* homologs of
 RT *Escherichia coli* cell division genes *ftsZ* and *ftsA*.";
 RL J. Bacteriol. 170:4855-4864(1988).
 RN [5]
 RN SEQUENCE OF 1410-1433 FROM N.A.
 RP STRAIN=168 / MARBURG;
 RX Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
 RA "Nucleotide sequence of the sporulation gene *spoIIA* from *Bacillus*
 RT *subtilis*.";
 RL Nucleic Acids Res. 18:657-657(1990).
 RN [6]
 RN SEQUENCE OF 195-219.
 RC STRAIN=NATTO 16;
 RA Kato T., Yamagata Y., Arai T., Ichishima E.;
 RT "Purification of a new extracellular 90-kDa serine proteinase with
 RT isoelectric point of 3.9 from *Bacillus subtilis* (natto) and
 RT elucidation of its distinct mode of action.";
 RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M29035; AAA62679.1; -
 DR EMBL; J05400; AAA83362.1; -
 DR EMBL; M2630; AAA22458.1; -
 DR EMBL; X17344; CAA35224.1; -
 DR EMBL; Z99111; CAB13403.1; -
 DR EMBL; Z99112; CAB13404.1; -
 DR F01; A35131; A35131.
 DR F01; A35750; A35750.
 DR F01; A36734; A36734.
 DR HSP; P00782; 288T.
 DR MEROPS; S08.017; -
 DR Subtilist; BG10233; bpr.
 DR InterPro; IPR002029; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASIN-ASP; FALSE_NEG.
 DR PROSITE; PS00137; SUBTILASIN-HIS; 1.
 DR PROSITE; PS00138; SUBTILASIN-SER; 1.
 KW Hydroxylase; Serine protease; Zymogen; Signal; Complete proteome.
 FT SIGNAL 1 30
 FT PROPEP 31 194 POTENTIAL.
 FT CHAIN 195 755 BACILLOPEPTIDASE F.
 FT PROPEP 756 1433 POTENTIAL.

FT ACT_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 274 274 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 452 452 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CONFLICT 219 219 T -> A (IN REF. 6).
 FT CONFLICT 393 393 A -> V (IN REF. 3).
 FT CONFLICT 829 834 KHONKA -> N (IN REF. 3).
 FT CONFLICT 836 841 QPOVLP -> RTRLYS (IN REF. 3).
 FT CONFLICT 844 852 AOVSVYTG -> FCRSRHKS (IN REF. 3).
 FT CONFLICT 853 1433 MISSING (IN REF. 3).
 SQ SEQUENCE 1433 AA; 1545.7 MW; 98DF6846897807C9 CRC64;
 Query Match 14.88; Score 510; DB 1; Length 1433;
 Best local similarity 26.48; Pred. No. 3.4e-21;
 Matches 210; Conservative 104; Mismatches 302; Indels 178; Gaps 32;
 QY 3 RLGAIVYLVGLL---AGIALAAPVKPVVRN-----NAVQKNYGLLTPGLFKKQV 53
 DB 8 RLSSVLTIVVISSLLPPGAAGASSKVTSPVKKELOSAESIQNKISSLKKS-FKKKEK 66
 QY 54 MNQNEVDIVMFGSYGDRDPAKVL-----RLMGAQVYKSYKIIPAVAVKKA----- 102
 DB 67 TTF-----LIKFDLANPEKAAKAAVKAASKLSAARTYQKRSVAVVSLKVTADESQ 120
 QY 103 RDLILI-----AGMIDTGY-----FGN 119
 DB 121 QDVLYKLTOKDKGNADQIH:YVYVNCIAVHASKEVMKVVQFPEVEKVLNPKRQLFKS 180
 QY 120 TRVSGTKFIQEDYKVOVDDATSYQCAQTVNNSLCYDGSQVVAIVDTGIDANHPDLKC 179
 DB 181 SPSFNKKAQKAIRATQGVFWNVQDIDAPKAM-ALGYDGTGTVAASIDTGVEMHHHPALKE 239
 QY 180 KVIG-----WYDAVNCRTSTPYDQOQGHVAGIVACTGVSNSQVIGVAPGA 225
 DB 240 KYRGYNPNPNPEPEENMNMW:DAVAGEASPYDLAGHGVHTGTMTVGSPEPDGTNQIGVAPGA 299
 QY 226 KLVGVKVLGADSGSVSTIIAGVDWVYQNKQYV-----IRVIN:SLSSQSSQSDTDS 278
 DB 300 KWTAVKASEDG-GTDADILEMSEWVLAPEAGNPHPEMAPDVVNNWSG---GGSGIDE 355
 QY 279 LQAVNNWADAG-IVVCV:MAGNS-----GPNTYTVGSPAAASKYITVGVADSNINIASF 331
 DB 356 WYRDMYNWAMRAADIPPEFSAGMQLFTPGCG--SIANPANYPESFATGATDINKLADF 413
 QY 332 SSRGPTADGLKPEVAVKVNLEAPRASGTSMTPTINDYTKASCTSMATPHSGVGCALI 391
 DB 414 SLQGPSYDEIKELISAP:YNI-----VPGQTYEDGWDGTSNAGPHSVAVANLL 465
 QY 392 LQHPSTWPKVKYKALIEPTDTPAKELADI---AYGAGRVNKKYKAYDDYAKLTPTGS 448
 DB 466 KQANASLYDEMEDITLNTA:PTFTDSTFPDPPNNGYGHGLVNAEDAV-----SA 514
 QY 449 VAD-KGSATHTFDVSG-----ATFVTATLYWDTGSSDIDLVDPNNGNEVDYSYATY 500
 DB 515 VTGGLKAEQVSVEGDDQ:LVYVQHEKYTEAVEGSLPLTLTRED---NYSVTSVKLSY 571
 QY 501 GPEKGYVNPRTA-----CTWTYKVSYKGAANYQDVVYSD---GSLSSGGCNPDPN 549
 DB 572 KLDQGEWTEITAKRISGDH:KGYVQAEIPDIKTKLSYKWMIHDFGHVSSQVDYVTK 631
 QY 550 PNPENPTPTTDTOT-----FSTSYNDY-WDTSPTFMVNSG---ATKPTGDLTPTDSY 598
 DB 632 PSITAGYKQDFETAPGGWVA:KGTNNWMCVPTGNTAASGEKGVYGTNLGNVA----- 686
 QY 599 NDLDDLVLDP-----NCNLP:ETSSNS-----YEHVEYANPAGTWTFL-VYAYSTY 645
 DB 687 NSANNLNMPPTKAPDSGSLPT:YKSKSWHNLEDDFYGVYVFLPEGEKNWQAGVYNGKTS 746
 QY 646 GWADYQLKAVYVYG 659
 DB 747 SMTDEIDLSAYKG 760
 RESULT 11

SUBT_BACSA STANDARD; PRT: 381 AA.
 AC P00763;
 DT 21-JUL-1986 (Rel. 01, Last sequence)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Subtilisin amylosacchariticus precursor (EC 3.4.21.62).
 GN APR.
 OS Bacillus subtilis var. amylosacchariticus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID:1483;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89008194; PubMed-3139650;
 RA Yoshimoto T., Oyama T., Honda T., Tone H., Takeshita T.,
 RA Kamiyama T., Tsuru D.;
 RT 'Cloning and expression of subtilisin amylosacchariticus gene';
 RL J. Biochem. 103:1060-1065(1988).
 RN [2]
 RP PARTIAL SEQUENCE.
 RX MEDLINE-72266687; PubMed-4560201;
 RA Markland F.S., Kurihara M., Smith E.L.;
 RT 'Subtilisin Amylosacchariticus. II. Isolation and sequence of the
 RT tryptic and cyanogen bromide peptides';
 RL J. Biol. Chem. 247:5619-5631(1972).
 RN [3]
 RP SEQUENCE OF 107-381.
 RX MEDLINE-72266688; PubMed-5055784;
 RA Kurihara M., Markland F.S., Smith E.L.;
 RT 'Subtilisin Amylosacchariticus. 3. Isolation and sequence of the
 RT chymotryptic peptides and the complete amino acid sequence';
 RL J. Biol. Chem. 247:5619-5631(1972).
 CC -|- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in P1. Hydrolyses peptide amides.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPOREULATION, AND MANY MUTATIONS WHICH BLOCK SPOREULATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPOREULATION.
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: D002164; BAA00186.1;
 DR PIR: A00971; SUBSS.
 DR PIR: A41448; A41448.
 DR HSSP: P04189; 1SCY.
 DR MEPROS: S08.001;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE ASP; 1.
 DR PROSITE: PS00137; SUBTILASE HIS; 1.
 DR PROSITE: PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 30
 FT PROPEP 31 106
 FT CHAIN 107 381
 FT ACT_SITE 138 138
 FT ACT_SITE 170 170
 FT ACT_SITE 327 327
 FT ACT_SITE 191 191
 FT CONFLICT 191 191

FT CONFLICT 365 365 N -> D (IN REF. 2).
 SQ SEQUENCE 381 AA; 39457 MW; 2251BADE22B4824F CRC64;
 Query Match 14.88; Score 509.5; DB 1; Length 381;
 Best Local Similarity 37.86; Pred. No. 7e-22;
 Matches 137; Conservative 54; Mismatches 124; Indels 49; Gaps 13;
 QY 78 VLRLMAQVQKYSKIIIFAVAKIKARDLLIAGIMDTGYFCNTRVSGIKFIQEDYKQVD 137
 DB 58 VISEKGGVKQKQKYNVAAATLDEKAVKEL-----KADPSVAVVEEDH-IAHE 105
 QY 138 DATSV-----SOIGADTYWASLGYDGGVVAVTVDGIDANHPDLKGVKGYDAVNGRST 193
 DB 106 YAQSPYGISOLKAPAL-HCGGTGNSWVKVAVIDSGIDSSHPDL--NYRGASFVPSSTN 162
 QY 194 PYDD-QCHGTHVAGIVAGTGVNSQYIGVAPAKLVGVKVLGADGSGSVSTIIAGVDVW 252
 DB 163 PYDGGSSHGHVAGTIAAL-H-SIGVLGVSPASLYAVKVLDDSTGSGQYSWIIINGIENAI 221
 QY 253 ONKDKYGRIVNLSLGSQSSDGTDSLQAVNANWDAGIVVCAAGNSGP--NTYTVGSP 310
 DB 222 SN-----NDVYNHSLG---CPSGSTALKTVVDKNVSSGIVVAANGHSGSSSTVGYP 274
 QY 311 AAASKVITVGAVDSNDNINPSSRGPTADGRUKPEVAPGVVDIIAPRASGTSMTGTPINDY 370
 DB 275 AKYPTIAVGVNNSNQRAJESSNGSELD-----VMAPGVSIQSTLPGGT----- 319
 QY 371 YTKASGTSNATPHYSGVGA--TQAHPSPDVKVTKALTETADIVAPKEIADIAGAGRVN 430
 DB 320 YGAYNGTSNATPHYAGAAAILSKHPTWTNAQVRDLSESTATYLG-----NSFYVKGGLIN 375
 QY 431 VYKA 434
 DB 376 VQAA 379

RESULT 12

SUBT_BACST STANDARD; PRT: 381 AA.
 AC 229142;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Subtilisin J precursor (EC 3.4.21.62).
 GN APRJ
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Geobacillus.
 OX NCBI_TaxID:1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCINB 10278 / M7C-1823;
 RX MEDLINE-92231936; PubMed-1567435;
 RA Jang J.S., Kang D.O., Chae M.J., Byun S.M.;
 RT Molecular cloning of a subtilisin J gene from Bacillus
 RT stearothermophilus and its expression in Bacillus subtilis.";
 RL Biochem. Biophys. Res. Commun. 184:277-282(1992).
 CC -|- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in P1. Hydrolyses peptide amides.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPOREULATION, AND MANY MUTATIONS WHICH BLOCK SPOREULATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPOREULATION.
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

[illegible]

us-09-841-553-1.rsp

OK NCBI_TaxID=1409;
RN [1]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
RA Kobayashi T., Ito S., Yamashita O.:
RT "Structure of a new alkaline serine protease (M-protease) from
RT Bacillus sp. KSM-K16.";
RL Acta Crystallogr. D 51:199-206(1995).
RN [2]
RN SEQUENCE OF 1-23, AND CHARACTERIZATION.
RX MEDLINE=9558832; PubMed=7632397;
RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
RA Koike K., Kawai S., Ito S.:
RT "Purification and properties of an alkaline protease from
RT alkalophilic Bacillus sp. KSM-K16.";
RL Appl. Microbiol. Biotechnol. 43:473-481(1995).
CC -|- SUBUNIT: MONOMER.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC PDB: 1MPT; 22-JUN-94.
DR MEROPS: S08.010; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Subtilase_S8; 1.
DR PRINTS: PRO0723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; 3D-structure.
FT ACT_SITE 32 32 CHARGE RELAY SYSTEM.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM.
FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
FT ACT_SITE 269 269 AA; 26723 NW; 7A03C86D534A1D07 CRC64;
SQ SEQUENCE 14.2k; Score 488.5; DB 1; Length 269;
Query Match 41.4%; Pred. No. 6.8e-21;
Best Local Similarity 43; Mismatches 97; Indels 33; Gaps 10;
Matches 122; Conservative 43;
QY 142 VSGIGADTVNNSLGYDGGVYVAVYDGTIDANHPDLKGVKGVLDVAVNGRSTPYDQGHG 201
Db 8 ISRVQAPAHNR-GLTSGCVKVAVLDTGI-STHPDL--NIRGGASFVPEGFSTQDGNHG 63
QY 202 THVAGTVAGTGSVNSQYIGVAPGAKLVGVKGVLDGSGSVSTIIAGVDVWVQNKDKYGR 261
Db 64 THVAGTIAALNN-SIGVLGVAPSAELIYKVLGVKGVLDGSGSVSSIAQGLEMAGNN---GMH 118
QY 262 VINLSLSSOSSDGTSLSQAYNNAWDAGIVVCVACNSGNPTTYVGSPPAAKVTYGA 321
Db 119 VANLSLGSPPSPS---ATLEQAVNSATSRGLVVAASGNSGAG--SISYEARYANAVGA 173
QY 322 VDSNDNTASFSSRGPTADGRLKPEVVPVAPVDIIAPRASGTSMTPTINDYITKASGTAT 381
Db 174 TDONNRASFQYGCGLD-----IVAPGVNVQSTYPGST-----YASLNGTSMAT 218
QY 382 PHYSGVGLIIQAHPSWTPDKVKTALITADIVAPKEADIAYAGRVNVYKAK 436
Db 219 PHVAGVAALVQKNPSNWNQIRNLKNTATGLGNTNL----YGSGLVNAEATR 269

Search completed: October 31, 2002, 13:30:07
Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:25:48 ; Search time 29.5 Seconds
(without alignments)
3864,531 Million cell updates/sec

Title: US-09-841-553-1
Perfect score: 3437
Sequence: 1 MKRLGVVLLVLLAGT.....YAYSTVGWADYQLKAVYVG-659

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0s
Maximum Match 100s
Listing first 45 summaries

Database : SPYREMBL_19:
1: sp.archaea:
2: sp.bacteria:
3: sp.fungi:
4: sp.human:
5: sp.invertebrate:
6: sp.mammal:
7: sp.mhc:
8: sp.organelle:
9: sp.phage:
10: sp.plant:
11: sp.rodent:
12: sp.virus:
13: sp.vertibrate:
14: sp.unclassified:
15: sp.rvirus:
16: sp.bacteriap:
17: sp.archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	699	20.3	1245	Q9RL54	Q9RL54 streptomyces
2	669.5	19.5	442	Q31788	Q31788 bacillus su
3	637	18.5	444	Q9K817	Q9K817 bacillus ha
4	636	18.5	1102	Q95684	Q95684 streptomyces
5	633	18.4	412	Q9AER6	Q9AER6 thermopla
6	623	18.1	1220	Q9L0A0	Q9L0A0 streptomyces
7	619.5	18.0	1253	Q9FC06	Q9FC06 streptomyces
8	590	17.2	1239	Q9FB24	Q9FB24 streptomyces
9	564.5	16.4	379	Q9FDF4	Q9FDF4 bacillus l
10	563.5	16.4	374	Q9F942	Q9F942 bacillus l
11	562.5	16.4	374	Q9F943	Q9F943 bacillus l
12	559.5	16.3	379	Q95321	Q95321 bacillus l
13	558.5	16.2	374	Q9F941	Q9F941 bacillus l
14	556.5	16.2	379	Q45299	Q45299 bacillus l
15	550.5	16.0	379	Q45300	Q45300 bacillus l
16	549	16.0	379	Q45301	Q45301 bacillus l

17	547	15.9	382	2	Q45522	Q45522 bacillus sp
18	541.5	15.8	378	2	Q45466	Q45466 bacillus sp
19	538	15.7	310	2	Q9FDF3	Q9FDF3 bacillus l
20	536.5	15.6	1398	1	Q9P9L1	Q9P9L1 pyrococcus
21	536	15.6	310	2	Q9FDF2	Q9FDF2 bacillus l
22	533.5	15.5	378	2	Q45684	Q45684 bacillus am
23	532	15.5	310	2	Q9F7C2	Q9F7C2 bacillus l
24	525	15.3	621	2	Q53401	Q53401 alteromonas
25	522.5	15.2	382	2	Q87655	Q87655 bacillus su
26	522	15.2	621	2	Q9F486	Q9F486 alteromonas
27	519	15.1	639	2	Q9JRG8	Q9JRG8 alteromonas
28	515.5	15.0	386	2	Q56365	Q56365 thermococcus
29	512.5	14.9	381	16	Q07613	Q07613 bacillus su
30	509	14.8	379	2	Q66153	Q66153 bacillus sp
31	505.5	14.7	275	2	Q9R7J4	Q9R7J4 bacillus su
32	500.5	14.6	275	2	Q9JL66	Q9JL66 bacillus su
33	497	14.5	757	16	Q9K6G6	Q9K6G6 bacillus ha
34	495	14.4	374	2	Q45523	Q45523 bacillus sp
35	495	14.4	824	2	Q45464	Q45464 bacillus sp
36	493.5	14.4	379	2	Q45467	Q45467 bacillus sp
37	489	14.2	1433	2	Q45616	Q45616 bacillus su
38	483	14.1	401	2	Q9AER8	Q9AER8 bacillus st
39	477.5	13.9	799	16	Q9KEM1	Q9KEM1 bacillus ha
40	476	13.8	422	1	Q97F5	Q97F5 pyrococcus
41	474	13.8	361	2	Q45521	Q45521 bacillus sp
42	473.5	13.8	535	16	Q9KVI8	Q9KVI8 vibrio chol
43	473.5	13.8	715	2	P70765	P70765 alteromonas
44	463.5	13.5	382	2	Q9KWR4	Q9KWR4 bacillus pu
45	461	13.4	627	16	Q9RUD0	Q9RUD0 deinococcus

ALIGNMENTS

RESULT 1
Q9RL54 PRELIMINARY: PRT: 1245 AA.
ID Q9RL54
AC Q9RL54:
DT 01-MAY-2000 (TREMBLrel. 12, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 13, Last annotation update)
DE PROBABLE SECRETED PEPTIDE35.
GN SCF51A.10.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser M.M., Denepaite D., Eichner A., Cullum J.,
RA Kinashi H., Hepwood D.,
RT "A set of ordered cosmid and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RMol. Microbiol. 21:77-91 (1996).
DR EMBL: AL215196; CAB5642.1;
DR HSP: P00782; ISUE.
DR MEROPS: S08.09A;
DR InterPro: IPR003137; P
DR InterPro: IPR000209; P
DR Pfam: PF02225; PA: 1;
DR Pfam: PF00082; Peptidase_30; 1.

DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 SQ SEQUENCE 1245 AA; 130895 MW; 74E92DB9CALDE60 CRC64;
 Query Match 20.3%; Score 699; DB 2; Length 1245;
 Best Local Similarity 41.1%; Pred. No. 6e-28;
 Matches 175; Conservative 67; Mismatches 134; Indels 50; Gaps 13;
 QY 72 RDAVAVKLELGAQV-----KYSYKIIIPAVAVKIKARDLLIAGMDTGYGNTYV 122
 DB 157 RDASVRELOSIDARAVRDEAELEGREWKLVPAAGGAKA-----A 197
 QY 123 SGIKETQEDYKVVQVDDATSVSQICADTVVNSLGYDGSQVVAIVDTGIDANHPDLKGVY 182
 DB 198 AATPRVWLDGRVAGLSRSTAQIGAPDVN-SAGYRGDGVKVAVLDTGADQSHPLAGRYA 256
 QY 183 GYDVAVGRESTPYDDOCHGTHYACIVAGTGSVN-SQYIGVAPGAKLVGVKVLGADSGSYV 241
 DB 257 AKO-FSSGTSNDFVGHGTHVASTVGGSAASGSGRGVAPARLLVGVKVLGDDGFGSE 315
 QY 242 STIIAGVQVYQNKDKYTGIRVINSLSGSSQSDGSDTSLSQAVNN-AMDAGIVVCAAGNS 300
 DB 316 SQVIAGMEWAQDQ----GADVNNSLGSGCATDGTDPMSQALNLSRRTGTLFVVAAGNE 371
 QY 301 G-PNTYVGSPPAAKSVITVCAVDSNDNIASFSSRCP-TADGRKLPVWAPGVYDIAPKA 358
 DB 372 GEGGRTVSGPAGDAALTVGAVDRDSDSLAPFSSRGPRLGDDAVKPDVTPAGVGIVAAHA 431
 QY 359 SGTSMGTPIYNTYKASGTSNATPHVSGVYCALILQAHPSPDKVKLTALTIETADIVAPKE 418
 DB 432 AGSAMGDPVDEHYTAASGTSNATPHVAGAAALLAQHRPDWTAQGLKALLISTAVTVDCOK 491
 QY 419 IADIAGVGRVNYVKAIDYAKLTFTGSVADKGSATHTFDVSSATFTVATLYWDGSS 476
 DB 492 VFE--OGGRIDVRAA----GLGAVTATGL-----VMGPTSTRDTPVTSRVRY-TNS 535
 QY 479 DIDLYL 484
 DB 540 DEDVTL 545
 RESULT 2
 O31788 ID O31788 PRELIMINARY; PRT; 442 AA.
 AC O31788
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE ALKALINE SERINE PROTEASE.
 GN APRX
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borliss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert B., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris H., Karamata D., Kusanara Y., Kieckhefer M., Klein R.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krohn S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,

Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Puzelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scenlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccogna E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis".
 RL Nature 390:249-256(1997)
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR ENBL; 299113; CAB13610.1;
 DR HSP; 099405; IMPT.
 DR MEROPS; S08 UPA;
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 442 AA; 47905 MW; B996446AE87BADF2 CRC64;
 Query Match 19.5%; Score 669.5; DB 16; Length 442;
 Best Local Similarity 46.0%; Pred. No. 5.2e-27;
 Matches 151; Conservative 51; Mismatches 103; Indels 23; Gaps 9;
 QY 127 FIQEDKVVQVDDATSVSQICADTVVNSLGYDGSQVVAIVDTGIDANHPDLKGVYWD 186
 DB 116 YLNREKALDITATEASH-AEYVRNGQTUTGKGTVAVVDGI-YPHDLEGRIGFAD 173
 QY 187 AVNGRSTPYDDOCHGTHYACIVAGTGSVNS-OYIGVAPGAKLVGVKVLGADSGSVSTII 245
 DB 174 MVNQTPEYDDNGHGHGTHCAGI-VSSGASSGQYRGAPAPENLIGVKNKOGSGTLADII 233
 QY 246 AGVDMVWQ-NKDK--YGRVNLNLGSS---QSSDCTDLSQAVNNWDAGIVVCAAGN 299
 DB 234 EGVEWCIOXNEDNPDEPIDMSLSGLGDALRYDIEQEDPLVRAVEEANSAGIVVCAAGN 293
 QY 300 SGPNVTYVGSPPAAKSVITVCAVDSN-----DNIAFSSRSGPTADGRKLPVWAPGV 352
 DB 294 SGFDSQTIASPGVSEKVTVAALDNDNTASDDDTVASFSSRGPTVYKGRPDILAFGVN 353
 QY 353 IIAPRASGT-----SNGPPIKCYTKASGTSNATPHVSGVYCALILQAHPSPDKVKVTAL 407
 DB 354 IISLRSFNSYIDKLOKSRVSGSYFTNSGTSNATPICAGTAALILQONPDLTPDVEKELL 413
 QY 408 IETADIVAPKEADIATYGAERVNVYKAI 435
 DB 414 KNGTD--KWKDEDPNIGAGVNAENSV 439
 RESULT 3
 O3KB7 ID O3KB7 PRELIMINARY; PRT; 444 AA.
 AC O3KB7
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 25, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 13, Last annotation update)
 DE INTRACELLULAR ALKALINE SERINE PROTEASE.
 GN APRX OR BH930.
 OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN NCBI_TaxID=86665;
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii P., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis".
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001513; BAB05649.1;
 DR HSSP: Q99405; 1MPT.
 DR MEROPS: S08_UPA;
 DR InterPro: IPR001128; Cyt_P450.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 444 AA; 48916 MW; 3805F81C53A212F2F CRC64;
 Query Match 18.5%; Score 637; DB 16; Length 444;
 Best Local Similarity 39.2%; Pred. No. 2.4e-25;
 Matches 146; Conservative 68; Mismatches 122; Indels 36; Gaps 10;
 QY 85 QVKSYSKIPAVAKIKARDL--LLIAGMIDTVEGNTVSGIKPTQEDYKVVQDDATSV 142
 DB 87 OLRSFSKIRCVSANTVPSLOEMLVCKIRKIY-----LNRVHALDTAVES 130
 QY 143 SQICADTVNWSLGDGSGVVAIVDTGDANHPDLKGVIGWYDAYNGRSTPYDDGSHGT 202
 DB 131 AQ-APEVIRNGETLTKNDVTIAVIDTGI-YPHEDLEGRIKAFYDFVNOREEPYDDCKHGT 188
 QY 203 YHAGIVAGTG-SVNSOYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDVAVQNDKY---258
 DB 189 HCAGDAAGNAGSSDGVYGPAPAEVNVGVKVLNKGQGSLESIMQCVENCIYNEHPDD 248
 QY 259 GIRVNLISLSS---QSSDGLSLQVANNADAGIYVVAAGNSPNTYTVGSPAAASK 315
 DB 249 PIHISLSCGQALPYENQEDPMWRIVEEAHNAAGITVVAAGNSGPDQTIASPGVSE 319
 QY 316 VITVGAVD-----SNDNIASFSSRGPTADGRKPEVWAPGVDTIAPRASGT-----SM 383
 DB 309 VITVGALEDDRTDREDDVAFSSRGPTIYGPKPDPILAPGVNIVSLSPNSFFDKTK 343
 QY 364 GTPINDIYTRASCTSNATPHVSGVGLILQHPSTWDPKVKTLIETADIAPREIADIA 423
 DB 369 GSRVGSHTWNSGTSNATPCVAGVALLQHPNLTPDEYKTELMEGTDRWDRD--PNY 436
 QY 424 YGAGRVVYKAI 435
 DB 427 YGAGYISAGAI 438
 RESULT 4
 P95684
 ID P95684 PRELIMINARY; PRT; 1102 AA.
 AC P95684
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SUBTILISIN-LIKE PROTEASE
 OS Streptomyces albobrosculus
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomyces
 OC NCBI_TaxID=1887;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-S-3253;
 RX MEDLINE=97144528; PubMed=3990295;
 RA Suzuki M., Tsuchi S., Yamada S., Kojima S., Miura K., Momose H.;
 RT "A novel member of the subtilisin-like protease family from
 RT Streptomyces albobrosculus".
 RL J. Bacteriol. 179:430-438(1997).
 DR EMBL: D83672; BAA12040.1;
 DR HSSP: P00782; 28BT.
 DR MEROPS: S08_UPA;
 DR InterPro: IPR003860; BNR
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00102; BNR; 2.
 DR PRINTS: PR00082; Peptidase_S8; 1.
 DR PROSITE: PS00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Protease.
 SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE559E CRC64;
 Query Match 18.5%; Score 636; DB 2; Length 1102;
 Best Local Similarity 32.2%; Pred. No. 8.9e-25;
 Matches 201; Conservative 72; Mismatches 219; Indels 132; Gaps 26;
 QY 118 GNTRVSGIKFTQED--YKVOVDATSVSOIGADTVNWSLGDGSGVVAIVDTGDIDANHP 175
 DB 176 GORTASGVAVNLQGVKASLD--TSYGGIGTPKANEA-GYDGKGVIAVLDITGVDTATHP 232
 QY 176 DLKGVIGWYDAYNGRSTPY--DDQGHGTHVAGIVAGTGSVN-SQYIGVAPGAKLVGVK 232
 DB 233 DLKGVTA---SKFTSAPTGVGVGHGTHVASIAAGTCAQSKGTGVGAPGAKLVGVK 289
 QY 233 LGADSGSVSTIIAGVDVAVQNDKYGIRVNLISLQSSQSDGTDLSQAVNN-ANDAGI 291
 DB 290 LDDAGFGDDSGILAGMEAAQA---GADIVNLSGGNDTPE-TDPLEAVYDKLSAEKGI 344
 QY 292 VCVVAAGNSGNTYTVGSPAAASKVITVCAVDSNDNIASFSSRGPTADGRKPEVWAPG 350
 DB 345 LFAIAGNEGFPQ--SIGDSADSLTVAGVDDKDLADFSSTGPRIGDGVAKPDLTAPG 402
 QY 351 VDIAPRASGTSMTPTINYY---YTKASGTSNATPHVSGVGLILQHPSTWDPKVKTL 407
 DB 403 VDIITASAKGNDIAKEGVKTVAGYMTISGTSNATPHVAGAAALLKQHPKEMKVAELKGL 462
 QY 408 -IETADIAPREIADIAVAGRVVYKAIY-----DD---YAKLT 444
 DB 463 TASTKD---CKYTFPQGVGVQVQDKATQTVIAEPVSLSGVQVQHPHADDKPYTKRLT 518
 QY 445 F---TGSAVDKGSATHT-----FDVSGATFYATLYMDTSSDIDLYDPNGN 490
 DB 519 YRNLGTEVDYTLKLTSTATLQKKAAPAGPFTLGASTLTVA---NGTASVDVATDRLGG 575
 QY 491 EVDYSYATYGFKEGVGVYNYFTACTWTKVWSY-----KGAANYQ-----529
 DB 576 AVDTYSAYVATGACQSYTAAVEREVESYNTLVKLDKRSKATANTMAYLSGLTGLG 635
 QY 530 -----VDVYSDGLS---(DSB)-----GNPNPNPNPNPTPTTD 559
 DB 636 KDRSYAPTEADGAVSVRVPAGGVYLDASVYLGADPETWRGADWLAOPKLDVFRNTTIVTD 695
 QY 560 TQ-----TFTGSVNDYMDTSDTFTMVNVSAYKITGDLTFTSYNDLDLYDPNGNL 612
 DB 696 ARKAKPKVTVPCAKAKAFASADYTIETNDSAVSYGNWLENYSGFSRAHL-----GQ 749
 QY 613 VDRSTSSNSHEVYANPAGTWT 636
 DB 750 ITNGTLSQNM-NTHFSNGA-ROVT 772
 RESULT 5
 Q9AER6
 ID Q9AER6 PRELIMINARY; PRT; 412 AA.

AC Q9AER6;
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE THERMICIN (SUBTILISIN-LIKE PROTEASE THERMICIN).
 OS Thermoanaerobacter yonsei.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoanaerobacter group; Thermoanaerobacter.
 OX NCBI_TaxID=111519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jang H.J., Kim B.C., Pyun Y.R., Kim Y.S.;
 RT "A novel subtilisin-like serine protease from Thermoanaerobacter
 yonseiensis AB-1: cloning, expression and biochemical properties."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Jang H.J., Kim B.C., Pyun Y.R., Kim Y.S.;
 RT "Subtilisin-like protease, thermicin, from Thermoanaerobacter
 yonsei."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY028704; AA027733.1;
 DR EMBL: AF056633; AA09366.1;
 DR HSSP: Q45670; IDBI.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Protease.
 SQ SEQUENCE 412 AA; 44503 MW; E3C6A0F81B1A1D47 CRC64;
 Query Match 18.4%; Score 633; DB 2; Length 412;
 Best Local Similarity 41.5%; Pred. No. 3.5e-25;
 Matches 154; Conservative 52; Mismatches 129; Indels 36; Gaps 11;
 QY 83 GAQVYSYKIPAVAVKIKARDLILLIAGMIDTGYFNTKRVSGIKFIQED--YKVVQDDAT 140
 DB 59 GGSIKFPIPLKNGWSVNLPCDKLHPAS-----INGIHFIADSLVKQLQHLIAT 107
 QY 141 SVSQIGADTVNNSLGSDGSGVYVAIVDTGIDANHPDL---KCKVICHYDAVNGSRFPYDD 197
 DB 108 ---QETASRKANDLYTGKGTITAFDGTGI-YPHPDFTKPNRIIAFYDVYNGKQFDD 163
 QY 198 QHGHTVAGVAGTGT-SVNSQYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDVYVQNKD 256
 DB 164 NGHGHVAGDAAGNGYASNGKYGKGVAPRANIVAVKVLDSYGRGSSDILAGMOWLLNFE 223
 QY 257 KYGIRVNLISLGSSQS-SDGTDLSQAVNNANDAGIVVCVAGNSGPNITYVGSPPAAASK 315
 DB 224 KYRIVLSUSIGETPALETFDPLRVGVDPLKNGIIVVAVAGNSGPNYNSITSPQTSRN 283
 QY 316 VITVGAVD-----SNDNIASFSSRGPTADGRKLPVYVAPGVDTIAPRASGTSMTGP-- 366
 DB 284 AITVGAVDKRPDIEDDEVAKFSGRG--GPVLYKPDVYVAPGVKIVSTASGRVYFPGADEP 341
 QY 367 -INDVYTKASGSMATPHVSGVGLILQARPSWTDPKVKATLIETADIVAPKEIADIAZG 425
 DB 342 MINKPVSATGSMATPHVAGVALLLEKNSRLTNVEIKNLIKTTATKI--NEAGLWTGQ 395
 QY 426 AGRVNYKAIK 436
 DB 400 SGMINTEEALK 410
 RESULT 6
 Q9LQAO
 ID Q9LQAO PRELIMINARY; PRT; 1220 AA.
 AC Q9LQAO;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PUTATIVE SECRETED PEPTIDASE.
 GN SCC24.17C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Brown S.P., Harris D.;
 RC STRAIN-A3(2);
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN-A3(2);
 RC Cerdano A.M., Parkhill J., Bartrell B.G., Rajandream M.A.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=8813436;
 RA Redenbach M., Kieser H.M., Penapaitte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1998).
 DR EMBL: AL163003; CAB66111.1;
 DR HSSP: Q99405; IMPT.
 DR MEROPS: S08.0PA;
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR00209; Peptidase_S8.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 SQ SEQUENCE 1220 AA; 127111 MW; B9BD0018A7F31674 CRC64;

Query Match 18.1%; Score 623; DB 2; Length 1220;
 Best Local Similarity 36.4%; Pred. No. 4.8e-24;
 Matches 172; Conservative 81; Mismatches 155; Indels 64; Gaps 20;
 QY 28 PVYRNAVQKKY---GILFELPKKQVNNQVDTVINEGSGDGRDRAVKVLRMG 84
 DB 105 PYLRDGLDERFDVGLALGGA-----DSETGELPLLVITG---KGVRAATPRGA 153
 QY 85 QVRYSTIIPAVAVKIKARDLILLIAGMIDTGYF---GNTRVSGIKFIQEDYKVVQDDAT 140
 DB 154 ERTSLPSVRGAABE-----ADKRAFMRPFRGAGVEGVWLDGRVTADMAE 201
 QY 141 SVSQIGADTVNNSLGSDGSGVYVAIVDTGIDANHPDLKGVKIVGYDAVNGSRFPYDDQGH 200
 DB 202 SNAQIGTPANFA-GLTGCHITVAIVDSGVADGHPDLAGRIASRSPFGEEVA-DREGH 259
 QY 201 GTHVAGVAGTGSV-NSQYIGVAPGAKLVGVKVLGADSGSVSTIIAGVDVYVQNKDYG 259
 DB 260 GTHVTGVCSSGAASDGRG-YVAPGATLVGKVLDDGCGFSESEIAGMEWAARDVD--- 316
 QY 260 IRVNLISLGSSQSDGTDLSQAVNN-AMDAIVVCVAGNSGPNITYVGSPPAAASKVIT 318
 DB 317 ADIVNSLSTSPSDGTDFANAVNTLSRETGALFVIAAGNTGAPS-SIGSPGAADAALT 375
 QY 319 VGAVSDNDNIASFSSRG-GRKLPVYVAPGVDTIAPR---ASGTSMTGTPINDVYTKA 374
 DB 376 VGAVSDAQAAQAFISAGPRNGDNALKPLDSAPGVGILAAARSLAEGSG-----QTSIM 428
 QY 375 SGTSMATPHVSGVGLILQARPSWTDPKVKATLIETADIVAPKEIADIAZG--GAGRVNY 432
 DB 429 DGTSMATPHIAGVALLAEED-DNSGARLKDALKST-----KELDSAYOLGAGRVSV 483
 QY 433 KAIKYDDYAKLFTGSGVADKGSNHTTVDVSGATFVTATLYWDGTSSDIDLVL 484
 DB 484 AAVGAD-----VITATGS-ADLAFYSNPTEADPE--VITVAV-TNSTDIDVEL 527

```

RESULT 7
ID Q9FC06 PRELIMINARY: PRT: 1253 AA.
AC Q9FC06;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE SECRETED PEPTIDASE.
GN SCB11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.;
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL391041; CAC01576.1;
DR HSP; Q99405; IMPT.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EEDB89 CRC64;

Query Match
Best Local Similarity 37.4%; Score 619.5; DB 2; Length 1253;
Matches 164; Conservative 67; Mismatches 160; Indels 47; Gaps 37;

QY 118 GNTVSGIKFTQEDYKQVDDATSVQIGADTVWNSLGYDGSYVAIVDTGIDAN4QDL 177
DB 199 GRLAGGAKYVLDGKVRATLSTTAQIGAPDWMSG-GNTGEGVGVAVLDTGVDAGHPDF 257
QY 178 KGVICWYDAVNGRSTPYDDQGHGTHVAGIVAGTGSVNSQY-IGVAPGAKLVGVYVQAD 236
DB 258 AGRIANTASFPVDDQVDT-DRNGHGHVASTVAGTGAASGGVGVAPGASLHIGKVDNS 316
QY 237 GSGSVSTTIAGVDVAVQVQNDKYGIRVINLSLGSQSSDGTCSLSQAVNNAM---DAGVYV 293
DB 317 GSGQSWYLAQENAVRDQH---AKIVSKSLGDS-PTDGTDLPLSEAVN-NLSAETGLIF 370
QY 294 CVAAGNSQNTYTVGSPAAAKSVITVGAVDN-----DNIAFSSSRGP-TADGRLKPEVKA 348
DB 371 VVAAGNSPEAYTVCTPAADAALTVGAVNGPGKGVLDADFSRGPVGNVNAKVPDLGA 430
QY 349 PGVDIAPRAGSTSMGTPTINDYTKRSTSNATPHVSGVALLQNHPSWTPDKVKYALI 403
DB 431 PGVGLAARSRYAPEG---EGAYOSLSGTSMTPHVGAALALAEHPMTQGBLKBALY 487
QY 409 ETADIVAPKEADIAAGRVNYYKAIKYDDYAKLTFTGSVADKGSATHTFDVSGATFVY 469
DB 488 GTT---AGTQRFSPFDAGSGRVDAVAAR-----STLLASG---DAFAQAHPYTPGQIVRN 514
QY 469 ATLYNWTGSSDI---DLVLYD-----PNGNEVDYSYTAAYGFEKGYVNPAGTWTYKVVSYK 513

```

```

Db 539 DVTYNSGPAPVALDLALSPA3LPEG-----LPTLSEAOVTVPAHGTSVGVITHL 589
QY 524 GAAN----YQVDWSDGS 537
DB 590 DAMEDNGAVATRLVASGA 607

RESULT 8
Q9FBZ4 PRELIMINARY: PRT: 1239 AA.
AC Q9FBZ4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE SECRETED PEPTIDASE.
GN SCB11.16C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.;
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL391041; CAC01588.1;
DR HSP; Q99405; IMPT.
DR MEROPS; S08.00A;
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
SQ SEQUENCE 1239 AA; 133505 MW; 8F5E9AC68EB81260A CRC64;

Query Match
Best Local Similarity 32.1%; Score 590; DB 2; Length 1239;
Matches 172; Conservative 74; Mismatches 192; Indels 98; Gaps 19;

QY 110 GMDTGFEGNTRVSGIKFTQEDYKQVDDATSVQIGADTVWNSLGYDGSYVAIVDTG 169
DB 196 GYRSGTDAAFR-GGVAVHEDGADPVADLADTAQIGAPRAWAG-GNTGCGVEYAVLDTG 253
QY 170 IDANHPDLKGVIGVNDVAVNGSTPYDDQGHGTHVAGIVAGTGSVNS-QYIGVAPGAKLV 228
DB 254 VDAHPDLADRIAQRSFPEVNTD-DRDGHGTHVASTIAGTGAASACKEGVARGARLS 312
QY 229 GVKVLGADSGSVSTTIAGVAVQVQNDKYGIRVINLSLGSQSSDGTCSLSQAVNN-AM 287
DB 313 IGKVLNDSRCRQISWTLANLHFAVERH---AKIVNSLGSQSSDGTCSLSRAVDRLSA 369
QY 288 DAGIVCYAAGNSQNTYTVGSPAAAKSVITVGAVDNNDNIAFSSSRGPTADGRLKPEVY 347
DB 370 QTALFVVAAGNG-EAGS-APGVATSAITVAGVADTDTLAPSSGQPRVDGALKAPEIT 428
QY 348 APGVDTIAPRAGSTSMGTPTINDYTKRSTSNATPHVSGVALLQNHPSWTPDKVKYAL 407

```

429 APGVGILANSSFAAGG---NGAYQSLSGTSMATPHVAGAAALAAARPDLSGSAK--- 492
QY 408 IETADIVAPKEADLAIY-----GAGRVNYKAIKYDDYAKLTFTGSAVKGSATFTFDVS 462
Db 483 ----DVLASSHRTPRYDAFGAGGRVDVAAVRAGVYASATAYAFGSSPGPVRRLVYIT 538
QY 463 GATFVATLWTGSSDIDLIDPGRNEVDYSYATYIGREKVGYNPNPAGTWTKVYV 522
Db 539 NITGAATL-----ELSVAAHAPB--GVFLSASRTVTPA--- 572
QY 523 KGAANTOVYVSDGSLSGSGGNNPNPNPNPTPTDTGTTGTYNDYHETSDTFTMNV 582
Db 573 HGTAD--VTLTIDGSGSAGG-----RAYSGIL-----ATDAARVA 608
QY 583 SGAKINGDLTFTDTSVNDLIDLYDPNGNLVDRSTSSNSYERVEYANPAPGTWTEL 638
Db 609 HTAYS-AGPYR-----HKLTVHFKDAGS-----NPVEGVFDLL 640

RESULT 9
Q9FDFA PRELIMINARY: PRT: 379 AA.
ID Q9FDFA4
AC Q9FDFA4
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE KERA.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=OMU 1411T;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
RT Bacillus licheniformis OMU 1411T".
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL: AF282893; AAC00492.1;
DR HSP: P00780; ICSE.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
SQ SEQUENCE 379 AA; 38904 MW; DG65F6A93EBC69CC CRC64;

Query Match 16.4%; Score 564.5; DB 2; Length 379;
Best Local Similarity 35.6%; Pred. No. 1.1e-21;
Matches 155; Conservative 61; Mismatches 144; Indels 75; Gaps 14;

QY 7 VVALVVLGALLAGTALAAPVTVVVRNNAVQKNYGLTFLGFKKVRMNNQEVDTVIMF 66
Db 11 MLTALMLVTFNAFSDSASAAQPAK-----NVEKDYIVGF 44
QY 67 GSYGDRRAVK--VLRMLGAQVYKSIIPAVAVIKAKDILLIAGMIDTGYFGNTRVSG 124
Db 45 KS-GVKTASVKKDIKESGGVYKQFRIINAAKAKLDKEALEEYKNDPDVAY----- 95
QY 125 IKFIQEDYKQVDDAT---SVSQIGADTVNNSLGDGSGVYVAIVDTGIDANHPDLGKV 181
Db 96 ---VEEDRYAHALAQTVPYGIPIKADKV-QAQQGPKANVAVLDTGICQASHPDL--NV 149
QY 182 IGMTDAVNGRSTPYDDOGRGTHVAGTVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGV 241
Db 96 ---VEEDRYAHALAQTVPYGIPIKADKV-QAQQGPKANVAVLDTGICQASHPDL--NV 149
QY 150 VGGASVFAGEAVNTDGNRGHTHAGTVAALDNTTG-VLGVAPVSLVAVKVLNNSGSGY 208
Db 242 STIAGVDVYVQNKQYIRVNLNLGSSQSSDGTSLSSQAVNNADAGTVCVAPRSG 301
QY 209 SGIVSGIETATN---GMDVYNSLGL---GFGSGTAMKQAVDNAYARGVYVVAAGSG 261
QY 302 P--NTYTVGSPAAASKVTVGAVDSNDNIASFSSRGPTADGRLKPEVAPGVDDIIAPRAS 359
Db 262 SSGNTNTIGYPKAYDSVIRGAVDSNRSRASFSSVG-----AELEVHAPGACVYSTP 315

360 QTSMTGTPINDYTKASGISMATPHVSCVGLILQAHPSWTPDKVKTKALITADIIVAPKEI 419
Db 316 ST-----YATLNGTSMSP-IVAGAAALILSKHPNLSASQVNRKLSSTATYLG---- 362
QY 420 ADIAYGAGRVNYKA 434
Db 363 SSFYIGKGLINVEAA 377

RESULT 10
Q9F942 PRELIMINARY: PRT: 374 AA.
ID Q9F942
AC Q9F942
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 10689;
RA Evans K.L., Crowder J.S., Miller E.S.;
RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
RT feathers".
RL Can. J. Microbiol. 46:1004-1011(2000).
DR ENBL: AF205190; AAG31027.1;
DR HSP: P00780; ICSE.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
KW Signal; Hydrolase.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 >374 SUBTILISIN.
FT NON_TER 374
SQ SEQUENCE 374 AA; 38390 MW; B61432A1B42C526E CRC64;

Query Match 16.4%; Score 563.5; DB 2; Length 374;
Best Local Similarity 36.1%; Pred. No. 1.2e-21;
Matches 156; Conservative 61; Mismatches 140; Indels 75; Gaps 15;

QY 7 VVALVVLGALLAGTALAAPVTVVVRNNAVQKNYGLTFLGFKKVRMNNQEVDTVIMF 66
Db 11 MLTALMLVTFNAFSDSASAAQPAK-----NVEKDYIVGF 44
QY 67 GSYGDRRAVK--VLRMLGAQVYKSIIPAVAVIKAKDILLIAGMIDTGYFGNTRVSG 124
Db 45 KS-GVKTASVKKDIKESGGVYKQFRIINAAKAKLDKEALEEYKNDPDVAY----- 95
QY 125 IKFIQEDYKQVDDAT---SVSQIGADTVNNSLGDGSGVYVAIVDTGIDANHPDLGKV 181
Db 96 ---VEEDRYAHALAQTVPYGIPIKADKV-QAQQGPKANVAVLDTGICQASHPDL--NV 149
QY 182 IGMTDAVNGRSTPYDDOGRGTHVAGTVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGV 241
Db 96 ---VEEDRYAHALAQTVPYGIPIKADKV-QAQQGPKANVAVLDTGICQASHPDL--NV 149
QY 150 VGGASVFAGEAVNTDGNRGHTHAGTVAALDNTTG-VLGVAPVSLVAVKVLNNSGSGY 208
Db 242 STIAGVDVYVQNKQYIRVNLNLGSSQSSDGTSLSSQAVNNADAGTVCVAPRSG 301
QY 209 SGIVSGIETATN---GMDVYNSLGL---GFGSGTAMKQAVDNAYARGVYVVAAGSG 261
QY 302 P--NTYTVGSPAAASKVTVGAVDSNDNIASFSSRGPTADGRLKPEVAPGVDDIIAPRAS 359
Db 262 SSGNTNTIGYPKAYDSVIRGAVDSNRSRASFSSVG-----AELEVHAPGACVYSTP 315
QY 360 QTSMTGTPINDYTKASGISMATPHVSCVGLILQAHPSWTPDKVKTKALITADIIVAPKEI 419

```

Db 308 GYSTPTPTTYAT-LNGTSMASPHVAGAAALILSKHPNLSASQVRNLSSTATYLG-...
Qy 420 ADIAYGACRVNY 431
: ||| |||
Db 363 SSFYKGLINV 374

RESULT 11
Q9F943
ID Q9F943 PRELIMINARY; PRT: 374 AA.
AC Q9F943;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
CN SURC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9526682; PubMed=7747965;
RA Lin X., Kelemen D.W., Miller E.S., Shih J.C.;
RT "Nucleotide sequence and expression of kera, the gene encoding a
RT keratinolytic protease of Bacillus licheniformis PWD-1.";
RL Appl. Environ. Microbiol. 61:1469-1474(1995).
DR EMBL: S78160; AB34259.1;
DR HSP: P00780; ICSE.
DR MEROPS: S08.001; -.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
SQ SEQUENCE 374 AA; 38923 MW; 507E2409F1ADFE03 CRC64;

Query Watch 16.4%; Score 562.5; DB 2; Length 374;
Best Local Similarity 39.0%; Pred. No. 1.3e-21;
Matches 149; Conservative 54; Mismatches 130; Indels 49; Gaps 14;

Qy 57 NOEDVTVMFGSYGDRDRAV--VLRMGAGVYKYSKIIPAVAVKIKARDLLIAGMDT 114
Db 35 NYEKDYIVGPKS-GVKTASVKKDIIESGKGVKQFRIINAAKAKLDKEALKEVNDPV 93
Qy 115 GYFGNTRVSGIKFTQEDYKVVQDDAT---SVSQIGADTVNLSGSDGSGVYVAIVDTGID 171
Db 94 AY-----VEEDHVAHALAQTVPYGIPLIKADKV-QAOGFKGANVAVLDTGIQ 141
Qy 172 ANHPOLKGVIGWTDVNGRSTPYDDQGHGTHVAGVAGTGYGAPCAKLVGVK 231
Db 142 ASHPOL--NVVGGASFVAGAYNTDNGRGTHTVAGTVAALDNTTG-VLGVAPSVSLYAVK 198
Qy 232 VLGDGSGSVSTIIAGVQVQKQYKIRVINLSGSSOSSDGTDSLQAVNNAWDAGI 291
Db 199 VLNSSGSGSYGIVSGIETATN---GMDVINSLG---GASGSTAMKQAVDNAYARGV 251
Qy 292 VYCVAGNSGP--NTYTVGSPAAASKVITVAVDSNDNIASFSSRGPTADGRKPEVVP 349
Db 252 VYVAAAGNSGSGNTNTIGYPAKYDSVIAGVDSNRSASFSSVG-----AELEVNP 305
Qy 350 GVDIAPRASGTSNGTPIINDYTKASGTSMTAPHSQVGAALIQAHPSWTPDKVKTAIE 409
Db 306 G-----AGVYSTPTNTYAT-LNGTSMASPHVAGAAALILSKHPNLSASQVRNLS 356
Qy 410 TADIVAPKEIADIAYGAGRVNY 431
Db 357 TATYLG-----SSFYKGLINVEAA 377

RESULT 13
Q9F941
ID Q9F941 PRELIMINARY; PRT: 374 AA.
AC Q9F941;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
CN SUBC.
OS Bacillus licheniformis.

```

```

RESULT 12
Q53521
ID Q53521 PRELIMINARY; PRT: 379 AA.
AC Q53521;
DT 01-NOV-1996 (TRENBLrel. 17, Created)
DT 01-NOV-1996 (TRENBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE KERATINASE.
CN KERA.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9526682; PubMed=7747965;
RA Lin X., Kelemen D.W., Miller E.S., Shih J.C.;
RT "Nucleotide sequence and expression of kera, the gene encoding a
RT keratinolytic protease of Bacillus licheniformis PWD-1.";
RL Appl. Environ. Microbiol. 61:1469-1474(1995).
DR EMBL: S78160; AB34259.1;
DR HSP: P00780; ICSE.
DR MEROPS: S08.001; -.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
SQ SEQUENCE 379 AA; 38923 MW; 507E2409F1ADFE03 CRC64;

Query Watch 16.3%; Score 559.5; DB 2; Length 379;
Best Local Similarity 39.7%; Pred. No. 1.9e-21;
Matches 149; Conservative 54; Mismatches 133; Indels 49; Gaps 14;

Qy 57 NOEDVTVMFGSYGDRDRAV--VLRMGAGVYKYSKIIPAVAVKIKARDLLIAGMDT 114
Db 35 NYEKDYIVGPKS-GVKTASVKKDIIESGKGVKQFRIINAAKAKLDKEALKEVNDPV 93
Qy 115 GYFGNTRVSGIKFTQEDYKVVQDDAT---SVSQIGADTVNLSGSDGSGVYVAIVDTGID 171
Db 94 AY-----VEEDHVAHALAQTVPYGIPLIKADKV-QAOGFKGANVAVLDTGIQ 141
Qy 172 ANHPOLKGVIGWTDVNGRSTPYDDQGHGTHVAGVAGTGYGAPCAKLVGVK 231
Db 142 ASHPOL--NVVGGASFVAGAYNTDNGRGTHTVAGTVAALDNTTG-VLGVAPSVSLYAVK 198
Qy 232 VLGDGSGSVSTIIAGVQVQKQYKIRVINLSGSSOSSDGTDSLQAVNNAWDAGI 291
Db 199 VLNSSGSGSYGIVSGIETATN---GMDVINSLG---GASGSTAMKQAVDNAYARGV 251
Qy 292 VYCVAGNSGP--NTYTVGSPAAASKVITVAVDSNDNIASFSSRGPTADGRKPEVVP 349
Db 252 VYVAAAGNSGSGNTNTIGYPAKYDSVIAGVDSNRSASFSSVG-----AELEVNP 305
Qy 350 GVDIAPRASGTSNGTPIINDYTKASGTSMTAPHSQVGAALIQAHPSWTPDKVKTAIE 409
Db 306 G-----AGVYSTPTNTYAT-LNGTSMASPHVAGAAALILSKHPNLSASQVRNLS 356
Qy 410 TADIVAPKEIADIAYGAGRVNY 431
Db 357 TATYLG-----SSFYKGLINVEAA 377

RESULT 13
Q9F941
ID Q9F941 PRELIMINARY; PRT: 374 AA.
AC Q9F941;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SUBTILISIN PRECURSOR (EC 3.4.21.62) (FRAGMENT).
CN SUBC.
OS Bacillus licheniformis.

```



```

[1]
RN  SEQUENCE FROM N.A.
RC  STRAIN=11594;
RA  Pan F.M., Chen S.T., Chiu S.H.;
RT  *Sequence characterization of the precursor of one mutant subtilisin
RT  from Bacillus licheniformis*;
RL  Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR  EMBL: X91260; CAA6266.1;
DR  HSP: P00780; ICSE.
DR  MEROPS: S08.001;
DR  InterPro: IPR000209; Peptidase_S8.
DR  Pfam: PF00082; Peptidase_S8; 1.
DR  PRINTS: PRO0723; SUBTILISIN_A;
DR  PROSITE: PS00136; SUBTILASE_ASP; 1.
DR  PROSITE: PS00137; SUBTILASE_HIS; 1.
DR  PROSITE: PS00138; SUBTILASE_SER; 1.
KW  SIGNAL.
FT  SIGNAL          1   30   POTENTIAL.
FT  CHAIN           107  379   POTENTIAL.
SQ  SEQUENCE        379 AA; 35082 NW; DFEFEC4F8EECA5B CRC64;

Query Match          16.2%; Score 556.5; DB 2; Length 379;
Best Local Similarity 38.4%; Pred No. 2.7e-21;
Matches 150; Conservative 56; Mismatches 130; Indels 61; Gaps

QY  57  NQEVDTVMFGSYGDRDRAYK-VLRMGAQVKYSKILIPAVAYKIKARDLLLIAGHIDT 114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   35  NYEKDYTVGFKS-GYKTAASKKDIKESGGKVDKQFRIIAAKAKLDKEAKVEKNDPDV 93
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  115  GY-----FQMTRVSGKRFQEDYKQVDDATSVSQIGADTVNLSGLYDGSVVYAI 165
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   94  AYVEEDHVHGHLGQTPVPGIGLPAKD-KVQAQ-----GFGKANVKVAV 135
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  166  VDTGIDANHPDLKGVICGVDAVNGSEIYPDQGGHGTIVAGTGVSNQSYIGVAPGA 225
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   136  LDTGQKASHPLL--NVGGASFPAGAYWTDGNGHGTIVAGTVAALDNTTG-VLGVPVS 192
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  226  KLVGKYLGAAGSGSVYTIAGVDVYVKKDKYGLIRVINISLGSOSSDGTDSLQAVNN 285
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   193  SLYAVKVLNSSGSGSYSGTVSGIEWTTN----GNDVINSLG---GASGGTAMKQAVDN 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  286  ANDAGIVYVVAAGNSCP--NTTYTGVGPAASKVITVGAVDSDNDNIASFSSKPTADGR 343
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   246  AYARGVYVVAAGNSGSGENTTIGYPAKCDSPVPGEDSNRSFSSVG-----AE 299
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  344  PTVAPGVVDIIPRACSTVSTINDYTKASCTSNATPHVSGVGLILQAHPSMTPKV 403
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   300  LEVMAP-----VSGVSTPINT-ITTLNGTSMASPHVAGTSALLISKHPNLSASQV 350
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  404  KTALETADIVAPKETADIANVAGRVNYYKA 434
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   351  RNRLSRTATLVG----SSPVTGSLINVEAA 377
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
Q45300 PRELIMINARY; PRG; 379 AA.
ID AC Q45300
AC Q45300;
DT 01-NOV-1996 (TREMBLrel. 61; Created)
DT 01-NOV-1996 (TREMBLrel. 01; Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19; Last annotation update)
DE SUBTILISIN CARLSBERG PRECURSOR.
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid:1402;
RN [1]
RC SEQUENCE FROM N.A.
RA Pan F.M., Chen S.T., Chiu S.H.;
RT *Sequence characterization of the precursor of one mutant subtilisin
RT from Bacillus licheniformis*;

```

Search completed: October 31, 2002, 13:31:14
Job time : 35.5 sec

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2002, 05:31:18 : Search time 3395.5 Seconds
(without alignments)
16944.853 Million cell updates/sec

Title: US-09-841-553-2

Perfect score: 1977

Sequence: 1 ATGAAGAGGTAGGTGCTGT.....AGGCCGTGCTACTACGGG 1977

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:

2: gb_hgt:

3: gb_in:

4: gb_ov:

5: gb_pat:

6: gb_ph:

7: gb_pl:

8: gb_pr:

9: gb_ro:

10: gb_sts:

11: gb_sy:

12: gb_un:

13: gb_vl:

14: gb_ba:

15: em_fun:

16: em_hum:

17: em_in:

18: em_mu:

19: em_mu:

20: em_mu:

21: em_mu:

22: em_mu:

23: em_mu:

24: em_mu:

25: em_mu:

26: em_mu:

27: em_mu:

28: em_mu:

29: em_mu:

30: em_mu:

31: em_mu:

32: em_mu:

33: em_mu:

34: em_mu:

35: em_mu:

36: em_mu:

37: em_mu:

38: em_mu:

39: em_mu:

40: em_mu:

41: em_mu:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1977	100.0	1977	AR201152	AR201152 Sequence
2	833.2	42.1	898	AR009704	AR009704 Sequence
3	753.6	38.1	12452	AE010265	AE010265 Pyococcus
4	750.2	37.9	1962	AR201155	AR201155 Sequence
5	560.4	28.3	1236	AR201146	AR201146 Sequence
6	532.8	26.9	554	AR009707	AR009707 Sequence
7	248.2	12.6	4252	SCS1A	AL121596 Streptomy
8	191.6	9.7	2809	AR202321	AR202321 Sequence
9	191.6	9.7	3900	AB007809	AB007809 Streptomy
10	191.2	9.7	3030	VCA429230	AJ429230 Volvox ca
11	190	9.6	2539	AR202322	AR202322 Sequence
12	182.4	9.2	3348	D83672	D83672 Streptomyce
13	179	9.1	14376	AE013049	AE013049 Thermococ
14	170.8	8.6	40899	SCC24	AL163003 Streptomy
15	149	7.5	32704	SC8A11	AL301041 Streptomy
16	147	7.4	87076	AC005918	AC005918 Homo sapi
17	145.2	7.3	158063	AP001046	AP001046 Homo sapi
18	145.2	7.3	340000	AP001751	AP001751 Homo sapi
19	142.8	7.2	169585	AC078821	AC078821 Homo sapi
20	126.2	6.4	39369	AC115681	AC115681 Dictyoste
21	125.4	6.3	87201	AC007623	AC007623 Homo sapi
22	124.4	6.3	17723	AE007152	AE007152 Mycobact
23	124.4	6.3	97683	AC116548	AC116548 Dictyoste
24	124	6.3	145	AR009706	AR009706 Sequence
25	123.4	6.2	167254	CNS05TDS	AL357093 Human chr
26	123.2	6.2	151705	AC117900	AC117900 Rattus no
27	120.6	6.1	176351	AC117007	AC117007 Rattus no
28	118.2	6.0	171260	AC026743	AC026743 Homo sapi
29	118	6.0	115833	AC124319	AC124319 Homo sapi
30	117.8	6.0	169163	AC115666	AC115666 Rattus no
31	117.8	6.0	181179	AL161725	AL161725 Human DNA
32	117.4	5.9	88203	AC097628	AC097628 Takifugu
33	116.8	5.9	185377	AC096312	AC096312 Rattus no
34	116.6	5.9	15311	AE007164	AE007164 Mycobact
35	116	5.9	53667	MTV016	AL021841 Mycobact
36	115.2	5.8	1311	AX413622	AX413622 Sequence
37	115.2	5.8	313450	AL596170	AL596170 Listeria
38	115.2	5.8	319630	AX413016	AX413016 Sequence
39	115.2	5.8	349988	AX417046	AX417046 Sequence
40	114.8	5.8	203	AF230413	AF230413 Kytococcu
41	114.6	5.8	150754	AC023491	AC023491 Homo sapi
42	113.4	5.7	1296	AX413619	AX413619 Sequence
43	113.2	5.7	15311	AE007164	AE007164 Mycobact
44	113.2	5.7	58930	AC098321	AC098321 Rattus no
45	113	5.7	116595	AP003816	AP003816 Oryza sat

ALIGNMENTS

RESULT 1
AR201152
LOCUS AR201152
DEFINITION Sequence 11 from Patent US 6358736.
ACCESSION AR201152
VERSION AR201152.1 GI:2031040
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1977;
AUTHORS Takakura.H., Morishita.M., Shimojo.T., Asada.K. and Kato.I.
TITLE Thermostable protease
JOURNAL Patent: US 6358736-A 11 19-MAR-2002;
FEATURES Location/Qualifiers

linear PAT 20-APR-2002

source 1. .1977
BASE COUNT 453 a 659 c 554 g 311 t
ORIGIN

Query Match 100.0%; Score 1977; DB 6; Length 1977;
Best Local Similarity 100.0%; Pred. No. 9, le-239;
Matches 1977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGAGTTAGTGTCTGCTGGGACTGTTCAAGAAAGTCCAGAGGATGAACCTGGAAACAGGAGTG 180
DB 1 ATGAAGAGTTAGTGTCTGCTGGGACTGTTCAAGAAAGTCCAGAGGATGAACCTGGAAACAGGAGTG 180
QY 61 GGCCTTGGGCGACCCGCTTAAACCGGTTGTCAGGAACACGCGGTTCAAGAGAACTAC 120
DB 61 GGCCTTGGGCGACCCGCTTAAACCGGTTGTCAGGAACACGCGGTTCAAGAGAACTAC 120
QY 121 GGAAGTCTGACCCGCGGACTGTTCAAGAAAGTCCAGAGGATGAACCTGGAAACAGGAGTG 180
DB 121 GGAAGTCTGACCCGCGGACTGTTCAAGAAAGTCCAGAGGATGAACCTGGAAACAGGAGTG 180
QY 181 GACRCCGTCATATGTTCCGGAGCTACGGAGACAGGAGGAGGCGGTTAAGGTACTGAGG 240
DB 181 GACRCCGTCATATGTTCCGGAGCTACGGAGACAGGAGGAGGCGGTTAAGGTACTGAGG 240
QY 241 CTCATGGGCGCCAGGTCAGTACTCCTCAAGATTAATCCCTGCTGCGGGTTAAATA 300
DB 241 CTCATGGGCGCCAGGTCAGTACTCCTCAAGATTAATCCCTGCTGCGGGTTAAATA 300
QY 301 AAGGCGAGGAGCTTCTGCTGATCGCGGCGATGATAGACGCGGTTACTTCGGTACACA 360
DB 301 AAGGCGAGGAGCTTCTGCTGATCGCGGCGATGATAGACGCGGTTACTTCGGTACACA 360
QY 361 AGGCTCTCGGCGATTAAGTTTCATACAGGAGGATTAACAGGTTGAGTGACGCGCACT 420
DB 361 AGGCTCTCGGCGATTAAGTTTCATACAGGAGGATTAACAGGTTGAGTGACGCGCACT 420
QY 421 TCCGTTCTCCAGATAGGCGCGGATACCGTCTGGAATCCCTCGCTACGACGGAAGCGGT 480
DB 421 TCCGTTCTCCAGATAGGCGCGGATACCGTCTGGAATCCCTCGCTACGACGGAAGCGGT 480
QY 481 GTGGTGGTTCATCGTATACGAGTATAGACGCGGATACCGGATCTGAAGGCAAG 540
DB 481 GTGGTGGTTCATCGTATACGAGTATAGACGCGGATACCGGATCTGAAGGCAAG 540
QY 541 GTCAATAGGCTGATGACGCGCTCAACGCGGAGTTCGACCGGCTACGATGACGAGGACAC 600
DB 541 GTCAATAGGCTGATGACGCGCTCAACGCGGAGTTCGACCGGCTACGATGACGAGGACAC 600
QY 601 GGAACCCAGTTCGCGGATCGTTGCGGAAACCGGAGGTTTAACTCCAGTACATAGCC 660
DB 601 GGAACCCAGTTCGCGGATCGTTGCGGAAACCGGAGGTTTAACTCCAGTACATAGCC 660
QY 661 GTCGCCCGCGGCGAGTGTGCGGCGTCAAGGTTCTCGGTTGCGGAGGTTGCGGAGG 720
DB 661 GTCGCCCGCGGCGAGTGTGCGGCGTCAAGGTTCTCGGTTGCGGAGGTTGCGGAGG 720
QY 721 GTCTCCACCATATCGCGGCTGTTGACTGGGTCGTCAGAAACAGGACAGTACGGGATA 780
DB 721 GTCTCCACCATATCGCGGCTGTTGACTGGGTCGTCAGAAACAGGACAGTACGGGATA 780
QY 781 AGGCTCATCAACCTCTCCCTCGGCTCCCGAGAGTCCGAGGAGGACGCTCCCTCAGT 840
DB 781 AGGCTCATCAACCTCTCCCTCGGCTCCCGAGAGTCCGAGGAGGACGCTCCCTCAGT 840
QY 841 CAGGCGCTCAACCAACGCTTGGGAGCGCGGATATAGTCTGGGTCGCGCGCGGCAACAGC 900
DB 841 CAGGCGCTCAACCAACGCTTGGGAGCGCGGATATAGTCTGGGTCGCGCGCGGCAACAGC 900
QY 901 GGGCGGACACCTTACACGCTCGGCTACCGCGCGCGGAGGAGGATTAACCGTGGT 960
DB 901 GGGCGGACACCTTACACGCTCGGCTACCGCGCGCGGAGGAGGATTAACCGTGGT 960

QY 961 GCAGTTGACAGCAACGACAAATCCGCCAGCTTCTCAGCAGGGGACCGACCGCGGACGGA 1020
DB 961 GCAGTTGACAGCAACGACAAATCCGCCAGCTTCTCAGCAGGGGACCGACCGCGGACGGA 1020
QY 1021 AGGCTCAAGCGCGGAAGTGTCTGCCCGCGCGCTTGACATCATAGCCCGCGCGGACGCGGA 1080
DB 1021 AGGCTCAAGCGCGGAAGTGTCTGCCCGCGCGCTTGACATCATAGCCCGCGCGGACGCGGA 1080
QY 1081 ACCAGCATGGGCGACCCCGGATTAACGACTACTACCAAGGCGCTCTGGAACCAAGCATGGCC 1140
DB 1081 ACCAGCATGGGCGACCCCGGATTAACGACTACTACCAAGGCGCTCTGGAACCAAGCATGGCC 1140
QY 1141 ACCCGGCAAGCTTTCGGGCGGTTGCGCGGCTCATCTCCAGGCCACCGAGCTGAGACCGG 1200
DB 1141 ACCCGGCAAGCTTTCGGGCGGTTGCGCGGCTCATCTCCAGGCCACCGAGCTGAGACCGG 1200
QY 1201 GACAAGGTGAGACCGCCCTCATCGAGACCGCGGACATAGTCCGCCCAAGGAGATAGCG 1260
DB 1201 GACAAGGTGAGACCGCCCTCATCGAGACCGCGGACATAGTCCGCCCAAGGAGATAGCG 1260
QY 1261 GACATCGGCTTACGCTGCGGCTAGGCTGAGCTTACAGGCGCATCAAGTACGACGACTAC 1320
DB 1261 GACATCGGCTTACGCTGCGGCTAGGCTGAGCTTACAGGCGCATCAAGTACGACGACTAC 1320
QY 1321 GCAAAGCTCACCTTACCGCGCTCCGTCGCGGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
DB 1321 GCAAAGCTCACCTTACCGCGCTCCGTCGCGGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
QY 1381 GTCAGGCGGCGGCTTCTGACCGCGGCTTCTACTGGGACAGCGGCTCGGAGGCGGATC 1440
DB 1381 GTCAGGCGGCGGCTTCTGACCGCGGCTTCTACTGGGACAGCGGCTCGGAGGCGGATC 1440
QY 1441 GACCTCTACCTTACGACCCCAACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1441 GACCTCTACCTTACGACCCCAACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
QY 1501 GGTTCGAGAGGCTGCGGCTACTAACAACCGGACCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
DB 1501 GGTTCGAGAGGCTGCGGCTACTAACAACCGGACCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
QY 1561 AGCTACAAGGGCGGGGGAATACAGGCTGAGGCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
DB 1561 AGCTACAAGGGCGGGGGAATACAGGCTGAGGCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
QY 1621 TCCGCGGCGGCGCAACCCGGAATCCAAACCCCAACCCGGAACCCCAACCCGGAACCCCAACCCG 1680
DB 1621 TCCGCGGCGGCGCAACCCGGAATCCAAACCCCAACCCGGAACCCCAACCCGGAACCCCAACCCG 1680
QY 1681 CAGACCTTCAACCGGTTCCGTTACGACTACTGGGACACCGGAGACCTTCACGATGAC 1740
DB 1681 CAGACCTTCAACCGGTTCCGTTACGACTACTGGGACACCGGAGACCTTCACGATGAC 1740
QY 1741 GTCACACGCGGTGCCCAACGATTAACCGGTGACCTGACCTTCGATACCTTCTACACGAC 1800
DB 1741 GTCACACGCGGTGCCCAACGATTAACCGGTGACCTGACCTTCGATACCTTCTACACGAC 1800
QY 1801 CTGACCTCTACTCTAGCAACCGGACCTGTTGACAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 1860
DB 1801 CTGACCTCTACTCTAGCAACCGGACCTGTTGACAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 1860
QY 1861 AGCTACGAGCAGCTGAGTACGCGGACCCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
DB 1861 AGCTACGAGCAGCTGAGTACGCGGACCCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
QY 1921 GCCTACAGCAGCTGAGGCTGGGCGGACTACGCTCAAGGCGGCTGCTACTACGCG 1977
DB 1921 GCCTACAGCAGCTGAGGCTGGGCGGACTACGCTCAAGGCGGCTGCTACTACGCG 1977

RESULT 2
AR009704
LOCUS
DEFINITION Sequence 7 from patient US 5756339.

linear PAT 04-DEC-1998

```

RESULT 3
LOCUS      AE010265/c
DEFINITION Pyrococcus furiosus DSM 3638, section 140 of 173 of the complete
            genome.
ACCESSION  AE010265
VERSION    AE010265
KEYWORDS   1 GI:18893813
SOURCE     Pyrococcus furiosus DSM 3638.
ORGANISM   Pyrococcus furiosus DSM 3638.
            Archaea; Euryarchaeota; Thermococci; Thermococcales;
            Thermococcaceae; Pyrococcus.
REFERENCE  1 (bases 1 to 12452)
            Maeder,D.L., Weiss,R.B., Dunn,D.M., Cherry,J.L., Gonzalez,J.M.,
            DiRuggiero,J. and Robb,F.T.
            Divergence of the hyperthermophilic archaea Pyrococcus furiosus and
            P. horikoshii inferred from complete genomic sequences
            Genetics 152 (4), 1399-1305 (1999)
            MEDLINE
            PUBMED 99359404
            10430560
            2 (bases 1 to 12452)
            Robb,F.T., Maeder,D.L., Brown,J.R., DiRuggiero,J., Stump,M.D.,
            Reh,R.K., Weiss,R.B. and Dunn,D.M.
            Genomic sequence of hyperthermophile, Pyrococcus furiosus:
            Implications for physiology and enzymology
            Meth. Enzymol. 330, 134-157 (2001)
            MEDLINE
            PUBMED 21079003
            11210495
            3 (bases 1 to 12452)
            Weiss,R.B., Dunn,D.M., Robb,F.T. and Brown,J.R.
            The complete sequence of the Pyrococcus furiosus genome
            Unpublished
            4 (bases 1 to 12452)
            Weiss,R.B.
            Direct Submission
            Submitted (12-FEB-2002) Human Genetics, University of Utah, 20
            South 2030 East, Salt Lake City, UT 84112, USA
FEATURES             Location/Qualifiers
     source            1..12452
                        /organism="Pyrococcus furiosus DSM 3638"
                        /strain="DSM 3638"
                        /db_xref="taxon:186497"
     gene              132..1010
                        /gene="Pfi657"
     CDS               132..1013
                        /note="Function Code: 13.1 Translation: Amino Acyl tRNA
                        Synthetases; (hisS)"
                        /codon_start=1
                        /product="Histidyl-tRNA synthetase"
                        /protein_id="AAL81781.1"
                        /db_xref="GI:10893814"
                        /translation="HWAKSMEYIKIAETANKLSIRQLMGYEVVPIIEEYSPET
                        KGLTKFAYNFLLSPDVTIRIKIQKQNAKVTIAEVLCNGDKGKQASIEGTG
                        KGLGVALESTAITSLSLESTGIFDIQSGIEVWKELSGYWPKTEALIKRNFETI
                        ENIPIDGSKNNKSELNFRGRRSGIEKLDITVIDSIEDPRVFDLGTVPRLPYVTDII
                        FEIYSPQGVKPIGGGGEVVRGKAIGFVFLNRLLSLIRPKDKARKVISEIDPKKAY
                        STAREWRLAIPVRL"
                        1010..1624
     gene              1010..1624

```

(5'-phosphoribosyl)-4-imidazolecarboxamide isomerase'

/protein_id="AA01786.1"
/db_xref="GI:18893819"
/translation="MKRIYIPDAIDMDGKVVRLKKKKDEKVVYGDPLKKAEMFSEY
DKHIVDLGAFSGKQMDIDYEKIRDVGVQVIGDFEDYESIKKAYDIGENVYII
GTALDISLEKRTSEFEGTVSLDSKGGKIFTKGMIEESLVEDAYKMLRKFVNR
VTVVDQDTLGTETKIKRWGDEEFTVAGGVATPTDIIQLAEIGFSGVITIGKALYDE
TISLEALAEVAKSVGKKNYSS"
4570..5370
/gene="p21663"
/note="p21663"
/note="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family; (hisf)"
/codon_start=1
/transl_table=11
/product="imidazoleglycerol-phosphate synthase, cyclase
subunit"
/protein_id="AA01787.1"
/db_xref="GI:18893820"
/translation="MKLSLKKLKLKLVARLIIAALDKGRVVGKIGFKFNIRA
GDPTLAKRYESEDTEVDLDTASYEKRTLLDLVVERVAEIIYPTVGGGRTVRE
EAREIRKAGDIFETAAVPEPSELSKIEADLIGSANLVADIAKRWNGKFEVETV
RKPQGDVAENAKRVEELGAGEILLSDTQGTGGEDILTAKVADVAVNTVIAASG
AKGPEHELVFIRIGSDAALAASIFHYGEYTVRELKMFLLAKNGIPVELD"
5345..6010
/gene="p21664"
/note="p21664"
/note="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family"
/codon_start=1
/transl_table=11
/product="phosphoribosyl-AMP cyclohydrolase /
phosphoribosyl-atp"
/protein_id="AA01788.1"
/db_xref="GI:18893821"
/translation="NEFQDLIRGHEHELEKLEOVNNEKNNGIVPIVQDEKGVLT
LAYNKAARJKTLEGTIANIYRSKGRIRKKEGVCNLTQVKEIKIDCDNDAALLYS
PKGACHTGYSCFYRNLGEPSEVLPIDYLSIKLESEIRIKRKTTPVWEGSYTKLR
KREGEKTYIKKFGSEALVLAEKREIRIIEVADLLHLLVLTITNOLISGEVNNELRE
RRK"
6007..7023
/gene="p21665"
/note="p21665"
/note="Function Code: 1.7 Amino Acid Biosynthesis:
Histidine family"
/codon_start=1
/transl_table=11
/product="putative histidinol-phosphate aminotransferase
(imidazole)"
/protein_id="AA01789.1"
/db_xref="GI:18893822"
/translation="MKWELINFPYRAVEGCVRTWLDKNESPYDLPLQKEELIELE
KRIENRPHITSDPLREALAEFTGLKKNENAGNSDELINLVKMKXKIVYTSV
ITGNSFFKALHIGIPVKDIPLEKEDTIDGERIAEGSKAALFASNPNTGNSQEN
EVLKVDGSRVITLDASEPSEGSFIPKISEYENLVILRTFKAFLAGICKCGNIN
NEKIDALYRILPPLNLSLTVAIKMLHDIYVKRKIKLVIRERIRREFIETYS
PSEANFLMKLDAYDYLAKGIVVRKLSGRLEGHIRTIVTIGKKWENDELINKKEFL
CRG"
7014..7730
/gene="p21666"
/note="p21666"
/note="Function Code: 16.1 Conserved Hypothetical"
/codon_start=1
/transl_table=11

gene
CDS

gene
CDS

gene
CDS

Query Match 38.1%; Score 753.6; DB 1; Length 12452;
Best Local Similarity 63.6%; Pred. No. le-85;
Matches 1269; Conservative 0; Mismatches 674; Indels 51; Gaps 6;

Db	780	CAAAGCACACCCGAGCTGGGCTCCAGACAAGTAAAAACAGCCCTCATATAGAAACTGCTGA	839
QY	1236	CATAGTCGCCCCCAAGGAGATAGCGGACATCGGCTACGGTGGGGTAGGGTGAACGCTCTA	1295
Db	840	TATCGTAAAGCCAGATGAAATAGCCGATATAGCTACGGTGCAGGTAGGGTTAATGCATA	899
QY	1296	CAAGGCATCAAGTACGACGACTACGCGCAAGCTCACCTTACCGGCTCCGTCGCCGACAA	1355
Db	900	CANGGCTATAACTACGATAACTATGCAAGCTAGTGTCACTGATATGTTGCCAACA	959
QY	1356	GGGAAGGCCACCCACACCTTCGAGCTCAGGGCGGCACTTCCTGTCAGCGCACCTCTTA	1415
Db	960	AGGCAGCAAACTCACCACTGGTTATTAGCGGAGCTTCGTTGCTAACTGCCACATATA	1019
QY	1416	CTGGGACACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCCCAACGGGAACGAGGT	1475
Db	1020	CTGGGACAAATGCCAATAGCGACCTTGATCTTTACTCTACGATCCCAATGGAACACAGGT	1079
QY	1476	TGACTACTCTACACCGCTACTACGGCTTCGAGAGGTCGGCTACTACACCCGACCGC	1535
Db	1080	TGACTACTCTACACCGCTACTATGATTCGAAGGTTGGTTATTACACCCCACTGA	1139
QY	1536	CGGAACCTGGACGGTCAAGGTCGTGACGTACAGGGGGGGGGAACACTACAGGTGAGCT	1595
Db	1140	TGGAACATGACCAATTAAGGTTGTAAGCTACAGCGGAAGTGCNAATCAAGTAGATGT	1199
QY	1596	GTCAGGACGGGAGCTCCAGCCAGTCGGG	1625
Db	1200	GTAAGTGATGGTTCCTTTACAGCCCTGG	1229
RESULT 6			
LOCUS AR009707			
DEFINITION Sequence 11 from patent US 5756339.			
ACCESSION AR009707			
VERSION AR009707.1 GI:3968312			
KEYWORDS			
SOURCE Unknown.			
ORGANISM Unknown.			
REFERENCE 1 (bases 1 to 564)			
AUTHORS Mita, M., Yamamoto, K., Morishita, M., Asada, K., Tsunashima, S. and Kato, I.			
TITLE Hyperthermostable protease gene			
JOURNAL Patent: US 5756339-A 11 28-MAY-1998.			
FEATURES			
source 1..564			
BASE COUNT 121 a 195 g 161 g 87 t			
ORIGIN			
Query Match 26.9%; Score 532.8; DB 6; Length 564;			
Best Local Similarity 97.95; Pred. No. 9.2e-58;			
Matches 543; Conservative 0; Mismatches 17; Indels 0; Gaps			
QY	598	CACGGAACCCACGTTCCGGGTATCGTTCGGGAACCGGCAAGCTTAACCTCCAGCTACATA	637
Db	4	CACGGAACCTCACGTGGCGGAGCAGTTCCGGGAACAGCGAGCTTAACCTCCAGCTACATA	63
QY	658	GGCGTCCGCCCCCGCGCGCAAGTTCGTTCGGCGCTCAAGGTTCTCGGTCCGACGGTTCGGGA	717
Db	64	GGCGTCCGCCCCCGCGCGCAAGTTCGTTCGGCGCTCAAGGTTCTCGGTCCGACGGTTCGGGA	123
QY	718	AGCGTCTCCACCATCATCCGGGGTGTTCAGCTGGGTCTGCAGAACAGGACAAAGTACGGG	777
Db	124	AGCGTCTCCACCATCATCCGGGGTGTTCAGCTGGGTCTGCAGAACAGGATAGTACGGG	183
QY	778	ATAAGGTCATCAACCTCTCCCTCCGGCTCTCCACAGCTCCGACGGAACGACTTCCCTC	837
Db	184	ATAAGGTCATCAACCTCTCTCCCTCCGGCTCTCCACAGCTCCGACGGAACGACTTCCCTC	243
QY	838	ACTCAGGCCGTCAACAACAGCCCTGGGACGCCGGTATAGTACTGCTGGCGCGCGGCAAC	897

```

|||||
Db 244 ACTCAGGCGCTCAACAACGCTGGGACGCGGTATAGTCTGCGTCCCGCGGCAAC 303
QY 898 ACGGGCGGCAACACTACACCGTCGCTCACCOCGCCCGCGGAGCGTCAATACCGTC 957
Db 304 ACGGGCGGCAACACTACACCGTCGCTCACCOCGCCCGCGGAGCGTCAATACCGTC 363
QY 958 GTGCGAGTTGACAGCAAGCAACGACACGTCGCGGAGTCTCCAGCAGGCGGACGGGAC 1017
Db 364 GTGCGAGTTGACAGCAAGCAACGACACGTCGCGGAGTCTCCAGCAGGCGGACGGGAC 423
QY 1018 GGAAGGCTCAAGCGGAGTCTGCGGCGCGCGGCTTGACATCATAGCCCGCGCCAGC 1077
Db 424 GGAAGGCTCAAGCGGAGTCTGCGGCGCGCGGCTTGACATCATAGCCCGCGCCAGC 483
QY 1078 GGAAGGCTCAAGCGGAGTCTGCGGCGCGCGGCTTGACATCATAGCCCGCGCCAGC 1137
Db 484 GGAAGGCTCAAGCGGAGTCTGCGGCGCGCGGCTTGACATCATAGCCCGCGCCAGC 543
QY 1138 GGCACCGCGGACGTTTCGGG 1157
Db 544 GGCACCGCGGACGTTTCGGG 563

RESULT 7
SC51A
LOCUS Streptomyces coelicolor cosmid 51A.
DEFINITION AL121596 AL645882
ACCESSION AL121596.2 GI:20520772
VERSION 1
KEYWORDS binding-protein dependent; aminotransferase; beta-glucosidase;
membrane protein; lacI family transcriptional regulator;
Mark family protein; membrane transport protein; hydrolase; integral
pyrrolidone-carboxylate peptidase; secreted peptidase; secreted
protein; secreted solute-binding lipoprotein; SIR2-like regulatory
protein; sugar phosphotransferase; tetr family transcriptional
regulator; transcriptional regulator; transmembrane efflux protein.
Streptomyces coelicolor A3(2).
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J.,
AUTHORS Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
PUBMED 8843436
REFERENCE 2 (bases 1 to 42527)
AUTHORS Murphy,L. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 42527)
AUTHORS Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
JOURNAL Direct Submission
TITLE Submitted (28-SEP-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On May 9, 2002 this sequence version replaced gi:6002208
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Wellcome Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/s.coelicolor/) CDS are
numbered using the following system eg SC7B7.01c. SC (S
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length

```

```

in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov/ftp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtc, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 51A overlaps
cosmid 776 on the AseI-P genomic restriction fragment.
FEATURES
source
1. .42527
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 51A"
/complement(13..2310)
/gene="SCF51A.01c"
/note="SCC0423"
/complement(13..2310)
/gene="SCF51A.01c"
/note="SCF51A.01c"
len: 765 aa. Similar to many Streptomyces Coelicolor
integral membrane proteins including: TR:Q33902 [EMBL;
M64683] ORF1-4 (ACT11) (711 aa), fasta scores opt: 291
z-score: 295.3 E(1): 4.3e-09 27.7% identity in 779 aa
overlap, TR:Q92577 [EMBL; AL035569] SC9D9.14 (748 aa),
fasta scores opt: 1021 z-score: 1028.4 E(1): 0.40.68
identity in 779 aa overlap and TR:O88022 [EMBL; AL031107]
SC5A7.16C (705 aa), fasta scores opt: 639 z-score: 645.0
E(1): 1.4e-24 37.2% identity in 764 aa overlap. Contains
multiple possible transmembrane hydrophobic domains."
/codon_start=1
/transl_table=11
/product="putative membrane transport protein"
/protein_id="CB55653.1"
/db_xref="GI:6002209"
/db_xref="SPTREMBL:O9RL63"
/translation="MATLLRLGCAFRFRHWTLAAAVLVGLVGLMSAGEAGDSDS
DMFEGIESQAFDLINERFPGABEGADARILVADPGROVTASEYRAIDTLVAEVA
DGPOVAAYSDPLDAATVSDGSTAYATVTEVDNVTAGRADLTAAVARTSGLT
VEAGGAVAGEPPAGIGFELGIVAAVLLITFGLAAGLPLITAVVGLTASIV
ALGALCMSTASGDLAMMIGIANGVDYALLVVSRYREAREAGIDAREAGLAVTAGS
AVFAGLTYYIALAGLSVIGVPSLTKGLAAGAVYAVLTITLVPLALCGFWPDALL
ARVYKGRSAPPARRERKGRSTPPARRLRCGPGGAAARTREAGASVWRARFQRRP
VPVLLGAVLLGALVAPALDLRIHMGDEAKPTSTERRAYDALADGEGFNGPLTV
VWDARADRRHSAVAASDRLATKATGVSVNAPPRNAKATVIONVATPTSPETE
ELVHTIDERRHATESATGATFEVTTALNDVAKMADALIPYLLVYVGLAFLLLV
VFGVLVPLKALGFLLSVSGFAGLVTFQNGHAGLGLGVETGPMSTFPLMVG
VFGVLMYQYTVISNRREAVHVGEPAAVYVTFGRHSGRVVTAAVVMTFVAGPMTG
GETILLRWGRLAVAVLEDAFVVRVAVYFVAVLVLLGERAWMLPRRLDLLPRVDEGE
ALGRRTPAPALAAVPTDRODTVRA"
2423..2428
2431..2922
/gene="SCF51A.02"
/note="SCC0424"
2431..2922
/gene="SCF51A.02"
/note="SCF51A.02"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF51A.02."

```

```

/protein_id="CAB56654.1"
/db_xref="GI:5002210"
/translation="MPGWTAGLPHLKKRPRGHGAGRDPDGRVSAVPAACQVEMCPG
RRGGRTAVATARGPTVRPGPQFGGPLEHRORGLIAGLGGRAPEPEAPROVAAA
RA"
RBS
gene
complement(3024..3974)
/notes="SCF51A.03c"
/notes="SC00425"
complement(3024..3974)
/notes="SCF51A.03c"
/notes="SCF51A.03c, unknown, len: 316 aa. Similar to
Streptomyces coelicolor TR:CAB463185 (EMBL: AL096743)
hypothetical 30.5 KO protein SC17.02C (281 aa), fasta
scores opt: 424 z-score: 490.6 E(): 5.7e-20 36.6% identity
in 292 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF51A.03c"
/protein_id="CAB56655.1"
/db_xref="GI:5002211"
/db_xref="SPTREMBL:O9RL61"
/translation="MTLSLTHVGDRTVYAHGAAGTELFVYVPEAEVFAKPKPLHP
VFLSAVYDYRNDNHNKGLQLTASHLSQNLWGGNTIVHGGRTALPVRYSMA
HVSFEIGADNADHVAIERLWHPHGLWAEERIEVRDVPDGTGSMFLWTSAVT
NRKAEPLRGSTGTRPGAGTGLFWRGPRAPRGRTVFTQASAAASGEPEAPADGPH
GSAARLYVGGHDGDRGHATLFEHAPSNHDSBGKTHPAHMFVRSDFPAVAPSWA
FHEELVLPGETLARAYRVVYADGANDRVARVAHLAHLAW"
complement(3979..3984)
4159..4161
4172..4402
/notes="SCF51A.04"
/notes="SC00426"
4172..4402
/notes="SCF51A.04"
/notes="SCF51A.04, unknown, len: 76 aa."
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF51A.04"
/protein_id="CAB56656.1"
/db_xref="GI:5002212"
/db_xref="SPTREMBL:O9RL60"
/translation="MEJPHLIELRRAVERADNRVSRNGENATVALVNSDTAALA
4475..4478
4483..5319
/notes="SCF51A.05"
/notes="SC00427"
4483..5319
/notes="SCF51A.05"
to many e.g. Pseudomonas fluorescens SW:PRXC_PSEF
(EMBL:AF031153) non-heme chloroperoxidase (EC 1.11.1.10)
(chloride peroxidase), fasta scores opt: 243 z-score:
282.1 E(): 2.4e-08 28.4% identity in 275 aa overlap and
many putative hydrolases: Streptomyces coelicolor
TR:CAB44393 (EMBL: AL078610) putative hydrolase SC135.34C
(269 aa), fasta scores opt: 259 z-score: 300.1 E():
2.3e-09 29.7% identity in 269 aa overlap. Containing a Pfam
match to entry PF00561 abhydrolase, alpha/beta hydrolase
fold."
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="CAB56657.1"
/db_xref="GI:5002213"
/db_xref="SPTREMBL:O9RL59"
/translation="MTERLAVDGTIAVETGSGPLIVLHGMCDSPAIVRTWPQVLY
AACHRVAAVDLGGCESSVDMPQSRATAGDLIALIRHLGGPAGVYGVHSVGGAAIT
AAARESLTAVELAPETRKQSVRLGDLRYRFRGNRLGLGATGSGPLNRSLTD
VATPGVPAHMSERIGRIEISLLRPGNRKAMGRSAPTADAGLGNRCPLVYVMG
TLPDPWADPHAEAGSAYVADLPSGLGRLEMGAGHYPHDPDPDEVALVATERSDA

```

```

misc_feature RA
4624..5395
/notes="SCF51A.05"
/notes="Pfam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 74.40, E-value 2.4e-18"
5335..5339
5342..5875
Query Match 12.6% Score 248.2; DB 1; Length 42527;
Best Local Similarity 54.8%; Pred. No. 1.5e-22;
Matches 659; Conservative 0; Mismatches 508; Indels 36; Gaps 7;
QY 439 GCGCATACCTCTCGAACCTCCCTCGCTACGACGAAGCGGTGTGTGCTGCCATCTC 498
DB 9906 GCGCCCCGCGACGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9965
QY 499 GATACGGGTATAGACGCTAACCTACCCGATCTGAAGCCAGCTCATAGGCTGTGATCAC 558
DB 9966 GACACCGGTGCGACGAGACCATCCGACCTGCGCGCGGTGCGCGC---GGCCAAG 10022
QY 559 GCGGTCAACGCGAGGTGCGACCCCTACCATGACACGAGGACGACGACGACGACG 618
DB 10023 GACTTCTCGGCGAGCTCGGCGACCAACGCTCTTCGGGCGCGGACCCACGCTCCCTCC 10082
QY 619 ATCGTTGCGGGAACCGCGACGCTTAACCTCCAGTAC---ATAGCGCTCGCCCCCGCGCG 675
DB 10083 ATCGTGGCGGACGCGCGCGCTCGCGCGGAGCGGTCGAGGAGTTCGCGCGCGCGCG 10142
QY 676 AAGCTCGTGGCGGCTCAAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 735
DB 10143 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10202
QY 736 GCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 795
DB 10203 GCGGCGCATGAGTGGGCGGCGGACGACGCGGCGCGA-----CGTGTCAACAATG 10250
QY 796 TCCCTCGGCTCTCCAGAGTCTCCGACGGAACGACGCTCCCTCAGTCAGCGCGCTCAACAA- 854
DB 10251 AGCTTCGCTCTCCGCGGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACG 10310
QY 855 ---CGCTCGGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 909
DB 10311 CTCAGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10370
QY 910 ACTTACACGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
DB 10371 CCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10430
QY 970 AGCAACGACAACTATCGGCGAGCTTCTCCAGCAGGAGGCGG---ACGCGGACGGAAGGCTC 1026
DB 10431 CGGAGAGATTCCCTCGCGCGCTTCTCCAGCGCGCGCGCGCGCTGCTGCTGCTGCTG 10490
QY 1027 AAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1086
DB 10491 AAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10950
QY 1087 ATGGGACCGCGGATTAACGACTACTACACGAGGCTCTGGAACGACGATGCGCACCGCG 1146
DB 10551 ATGGGAGCGCGGTGCGAGGAGGCTACACCGCGCGCTCGGCGGACCTCGATGCGCGCGCG 10610
QY 1147 CAGCTTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1206
DB 10611 CAGCTGCGCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10670
QY 1207 GTGAAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1266
DB 10671 TTGAAGACCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10730
QY 1267 GCTTACGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
DB 10731 GCGCGCGCGGAGTTCGACGCTGCGCGCGGCGGAGTGGGCGCGCTGACCGCGCGGCGACA 10790
QY 1318 TACGCGAAGCTCACCTTC/ACGCGCTCGCTCGCGGACGAGGAGGAGCGCGCGCGCGCGCTTC 1377

```

Db	1164	CGGCGCCCGCCAGTTCACAGGCGGTGCGACCGCGCCGCGATCCTCCACG	1223
Qy	689	TCAAGGTTCTCGGTCCGGTTCAGTTCGGGAAGCGTCTCCACCATCATCGCGGGTGTGACT	748
Db	1224	GCAAGGTCTCTGACGACTCTCGGTTCGGCGGACGACGCCGCCATCTCTCGCGCGATGAGGT	1283
Qy	749	GGGTGTCCAGAACAGGACACAGTACGGGATAAAGGTCATCAACCTCTCCCTCGGGTCTCT	808
Db	1284	GGG-----CGGTCCGCGAGGGCGCGAGTCTCAACATGAGCCTCGGCGGCGCA	1331
Qy	809	CCCAGAGCTCCGACGGAAACGACTCCCTCACTAGTCAGGCGGTCTCAACAACCCCTGGGAGCGCG	868
Db	1332	TGGACACACCGGAGACCGACCTCGCTGGAGCGGGCGTTCGCAAGTGTCTCGCGCGAGAAGG	1391
Qy	869	GTATAGTAGTCGTGCGTCCGCGCGCAACAGCGGGCGCGAACAACCTTACACCTGTGGGTCTAC	928
Db	1392	GGTCTCTGTTCCGCATCGGCGCGCGGCAAGAGGGCCG-----GAGTCGATCGTTGCG	1445
Qy	929	CGCGCGCGCGAGCAAGGTATTAACCGTCGGTGTGAGTTGACAGCAACAGCAACATCGCA	988
Db	1446	CGCGCAGCGGACGCCGCTCACTGTCGGCGCGCTGACGACGAAGGACAAAGCTCGCGC	1505
Qy	989	GCTTCTTCACAGGGGACCCGAC---CGCCGACGGAAGGCTCAAGCGCGGAAGTGTGTGCC	1045
Db	1506	AGTTCTCTCTCAACGCGCTCGGCGTGGCGACGCGCGCATCAAGCGGAGCACTCACGGTCT	1555

DB	1506	AGTTCTCTCTCCACCGGCCCCGCTGGGACAGCGGCGCATCAAGCGGCGCATCGCGCT	1100
QY	1046	CGGGGGTTGACATCATAGCCGCCCGCGCAGCGGACACAGCATATGGGCACCCCGAT-----	1100
DB	1566	CGGGGTGGACATCATCGGCCGCTTCGGCGGAGGGCAACGACATCATGGCCAGAGGTCGGTG	1629
QY	1101	-----AAGACGACTACTACACCAAGCGCTCTGGGAACCAAGCATGGCCACCCCGCACGTTTCGG	1155
DB	1626	AGGACACGGCGCGCTACATGACCATCTCCGGCAGCTGATGGCAGCCGCGACGTCGCGG	1688
QY	1157	CGGTTGGCGGGTCACTCTCCAGGCCACCCCGAGTGTGACCCCGCGACAAAGTGAAGACCG	1211
DB	1686	GGCGGGCGGCTCTCTGAAGCAGCAGCACCCGCACTGGACCTCCGCCGAACTGAAGGGCG	1747
QY	1217	CCCTCATCGAGACCGCGGACATATGTGCCCCCAGAGATAGCGGACATCGCTTACGGTC	1279
DB	1746	CGCTACCGGCTCCACAAAGGCGGC-----AAGTACACCCCGTTCCAGGACGGGTT	1799
QY	1277	CGGGTAGGGTGAACGCTCTACAAGGCCATCAAGTACAGGACTACGCCAGTCACTCTCA	1331
DB	1797	CGGGCGGATCCACGGCCGACAAAGCGCTCCAGCAGACCGTGTATCGCCACCCCGCTTCGG	1851
QY	1337	CGGGTTCGGTCGCGACAAAGGGAAGCGCCACCCACCACTTCGACGTCAGC	1386
DB	1957	TCAGCTCTCGCGCTCCAGCTAGTGGCGGCACACCGCAGCGAGCGCGGTCAACC	1906

DB	1566	CGGGGTGGGACATCAGGCGCCCTCTGGGAGGAGGAGGACACATCATCGCGCGAGGACG	11566
QY	1101	----AAACGACTACTACACCAAGGCGCTCTGGAAACGACGCTGGCCACCCGCGACGTTTCGG	11567
DB	1626	AGGGACCGGCGCGCTACATGATCACCATTCTCCGGGACGCTGATGGCGACCCCGGACGTCGGGG	1687
QY	1157	GGGTGGCGCGCTCATCTCTCAGCGCCACCCGAGCTGGACGCCCGCGGACGAAGTGTGAAGCGG	1211
DB	1686	GGCGGGCGGCTCTGTGAGCAGCAGCACCGCGACTCTCCGCGAACTGAAGGGCG	1747
QY	1217	CCCTCATCGAGACCGCGGCAGCATATGTCCGCCCCCAAGGATAGTGGGACATACGCTACCGTGTG	1277
DB	1746	CGCTCACCGGGTCTCAACAAAGGCGGC-----AAATACACCCCGTTCGAGCAGGGTT	1797
QY	1277	CGGGTAGGGTGAACGTCTTACAAGGCCATCAAGTAGCAGGACTTAGCCAGTCACTTCA	1337
DB	1797	CGGGCGGATCAGGCGCGCAAGGCGTCCAGCAGACGGTGTACGCGACCGCGGTCTCGG	1857
QY	1337	CGGGTTCGTCCGGACAAAGGGAAGCGCCACCCACACCTTCGACGTCAAG	1386
DB	1957	TTAGCTCTCGCGTCTCCAGATGTGGCGGCACACCGACGAGCGAGCGCGGTCAAC	1906

RESULT 9	
AB007809	3900 bp DNA linear BCT 07-JUN
LOCUS	
DEFINITION	Streptomyces viridisporus dhpA gene for 1,4-dihydropyridine enantioselective esterase, complete cds.
ACCESSION	AB007809
VERSION	AB007809.1 GI:21332743
KEYWORDS	
SOURCE	Streptomyces viridisporus (strain:A-914) DNA.
ORGANISM	Streptomyces viridisporus
	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomy-
REFERENCE	1
AUTHORS	Ariasa, A., Matsushita, M., Kasehara, T., Dobashi, K., Yoshioka, T., Yamada, S., Komase, H. and Teguchi, S.
TITLE	Streptomyces Serine Protease (DHP-A) as a New Biocatalyst Capable of Forming Chiral Intermediates of 1,4-Dihydropyridine Calcium Antagonists
JOURNAL	Appl. Environ. Microbiol. 68 (6), 2715-2725 (2002)
MEDLINE	22034940
PUBMED	12039725
REFERENCE	2 (bases 1 to 3900)

AUTHORS TITLE JOURNAL

Arizawa, A.
Direct Submission
Submitted (07-Oct-1997) Akira Arizawa, Bioresource Laboratories,
Mitsubishi Chemical Corporation, 4-9-1 Johnan, Fujisawa, Kanagawa 251-0057,
Japan (E-mail: arizawa-a@mercian.co.jp, Tel. 81-466-35-1511,
Fax: 81-466-35-1530)

FEATURES

Location/Qualifiers

1..3900
/organism="Streptomyces viridosporus"
/strain="A-914"
/db_xref="taxon:67581"

-35_signal

-10_signal

gene

338..3655

/gene="dhpA"

338..3655

/gene="dhpA"

/function="forming chiral intermediates of

1,4-dihydropyridine calcium antagonists"

/function="serine protease"

/note="gene product (DHP-A) has both serine protease

activity and enantioselective esterase activity"

/citation=[1]

/codon_start=1

/transl_table=11

/product="1,4-dihydropyridine enantioselective esterase"

/protein_id="AAC00400.1"

/db_xref="GI:21322744"

/translation="MKRAVATVATAAVALAAGMTGPAASGEHTAAQOPAKASA

KTSLSKATHTLITGDRVADTVAGVLEFRAEGREHIPVQIRKVDHILVLPADAA

RIVASGDLRLDELTELGAATRNQKQGLKIVIGYGAAAKAEVRAEGLRRLT

TSLNADVRPHEDASELDWATNGORTASGIAIIVLDGVRRAALDTSVQICAPKAM

SAGDQKGYAVLVDGTVSHDPLKGRVTAASKNFTAAPGADKVGHGTAVASIAAGT

QAGQKGYKAVPGRALVNGVLDSDGFGDDSGILAGMEWAAGADVWNSLGGHDT

PETLPDLAAVDLSAEKGVLPFAIRAGNEGPESIPGSADALVAVDDDKLADES

STGRPLGDGAIKPDVAPGVDTASAGNDIGOEVEGEPAGYMTISGTSNHTVYAG

AAALKKQHPDWTSAELKALGTSTKGGKYTPPEQSGRIQDRALQTVIADPVSVS

FCVQNPHTDDBPYTKQLYNLGTQDTVLKLSIATDPKGAAPAGFTVLGATTVY

PAGSASVDMATRLUGTVDCNATVAVTGGGTQVTAATAVREVEVDVTVRHIG

RDGAPTEHLTDLIAGLGRSGRGAPATDTATLRLPKGTVLVDSDIAKDFTLKGG

IDMLVPLSVTKDITLIDATTTAAADITVPDPKAPLSATIGYDTAGIGIASM

ESFADVRMAHLEAPFGIRIOTWNGQMTKSTEDYATSAKVKIKQKVRHFRAGE

LATARKLELAAGATGGVYTFGLPDPDFVEGAPAEQKLPGRSLKLSAEKRNFSFD

LEQRAQDAAGHPVLDVATVLEGPOTFKAGKVTFRFTAVFGPRLSADVGLERDGM

IVGLPFDADGKHAGSSDFTSATISLYRNKIKGSHSDPLLGDDBEFTVAGDASKU

TTSVKRSARVAATASTRIDASKMTFRSKTGTQLPASTVRFQAVIGLDSKVPAGKAT

PPVVEAGAKGNKLSLVSYSDSGTKMTKTTVTKGKITVNPAPAKKASIPRAKTD

KKGWTSLSYNNAYHGK"

338..349

/gene="dhpA"

/evidence="experimental"

3687..3723

/function="potential transcription terminator"

BASE COUNT 675 a 1471 c 1304 g 450 t

ORIGIN

Query Match

Best Local Similarity 9.78; Score 191.6; DB 1; Length 3900;

Matches 593; Conservative 0; Mismatches 439; Indels 48; Gaps 8;

QY 332 TGATACACAGCGGTTACTTCGTTACACAGAGGCTCTCGGCATATAAAGTTTCATACAGGAGG 391

DB 870 TGTGGGACCGCTTACACAGGCGAGCGGCGGCTCGGCATCGCCACGCTCGGCTGG 929

QY 392 ATTACAAAGTTTCAGGTGACGAGCGGCTTCGCTCCGATAGAGGCGGCGATACCGTCT 451

DB 930 ACCGGGTTCGGACGGCGGCTCGACACGCTCGCTCGGCAGATCGGCGCCCGCCACAGGGGT 989

QY 452 GGAATCTCCCTCGGTACGAGGAGCGGTGCTGGTGGTTCGATCGTGCATACCGGTATAG 511

DB 990 GG---TCCCGCGGTACAGCGGAGGGCGGTGAAGATCGCGCTCGGTGGACACCGGTGTCG 1046

QY 512 ACCGGACACCGCGATCTGAAGGGGCAAGGTCATAGGCTGTACGACCGCGGTCAACGGCA 571

Db 1047 ACACAGGATCCCGGAGCTTGAAGGGCCGGGTGA---CCGGTCCAAAGAACTTCACCGCGG 1103
QY 572 GTTCGACCCCTAGCATGATGACAGGAGACGGAACCCACGTTTGGGGTATCTGTTGCGGAA 631
Db 1104 CCGCGGGCGCGGCGGACAAAGTGGGCCACGCGACCCACGTCGATCGGCGGGGCA 1163
QY 632 CCGGAGC---GTTAACTCCAGTACATAGGCTTCCGCCCGGCGGGAAGCTCTGCGGG 688
Db 1164 CCGGCGCCCACTCCNAGCGCAAGTCAAGGGCTGCGCACCCCGCGCGGATCTCTCAACG 1223
QY 689 TCAAGGTTCTCGGTGCGGACGCTTCGGAAGCGTCTCCACCATCATCGCGGGGTGTGACT 748
Db 1224 GCAAGTCTCTCGACGACTCTGCTTTTCGCGGACGACTCCGGCATCTCGCGCGCATGAGT 1283
QY 749 GGTGTCTCCAGAACAGCAAGTACAGGGGATAAGGCTCATCAACCTCTCCCTCGGCTCT 808
Db 1284 GGG-----CGCGCGGCGAGGGCGGAGCTGCTCAACATGAGCTTGGCGGCA 1331
QY 809 CCCAGAGTCCGAGGGAACGGAAGTCTCTAGTCAGGCGGCTCAACAGCGCTGGGACGCG 868
Db 1332 TGGACACACCGGAGACGACCGCGCTGGAGCGGCGTCAAGAGCTGTCCGCGGAGAGG 1391
QY 869 GTATAGTGTCTGCTGCGCGCGGCAACAGCGGCGGCAACACTACACCGTCTCGGCTCAC 928
Db 1392 GCGTCTCTTTCGCGATCGCGCGCGGCAACAGAGGCGCG-----GAGTCGATCGTTCG 1445
QY 929 CCGCGCGCGGAGCAAGGTCATTAACCGTCTGCGTGGTGGTGGAGCAAGCAACATCGCCA 988
Db 1446 CCGGCGAGCGGAGCGCGCGCTCACGCTGGCGCGCTCGACAGCAAGCAAGCTCCCGG 1505
QY 989 GCTTCTCAGCAGGGAAGCGGAC---CGCGGAGCGGAGGCTCAAGCGGGAAGTCTGCTCCC 1045
Db 1506 ACTTCTCTCAGCGGCGCGCGCTCGCGGAGCGGCGGCTCAAGCGGAGCTCACCGCTC 1565
QY 1046 CCGCGGTTGATCATATAGCCCGCGCGGCGGCAACAGCAAGCAAGCTACACCGTCTCG 1100
Db 1566 CCGCGGTTGATCATAGCGCGCTCGCGGAGGCAACGACATCGCGCGAGGCTCGGTG 1625
QY 1101 ----AAGCGACTACTACAAAGGCTCTGGAACCAAGCATCGCCACCGGCTTCGG 1156
Db 1626 AGGAGCAGCGCGGCTACATGACCATCTCGCGGAGCTGCGATGGGAGCCCGGCGCGG 1685
QY 1157 GCGTGGCGGCTATCTCTCCAGCGGCGGAGCTGGAGCCCGGCAAGGTGAAGACCG 1216
Db 1686 GCGCGCGCGGCTCTCTGAAGCAGCAGCAGCAGCGGCTCGCGGAGCTCGCGGAGGCG 1745
QY 1217 CCCTCATCGAGAGCGCGCATAGTTCGCCCGCAAGGAGATAGCGACATCGGCTACGGTG 1276
Db 1746 CGCTCAGCGGCTCCACCAAGGCGG-----AAGTACACCGGTTTCGAGAGGTT 1796
QY 1277 CCGGTAGGTGAGCTCTCAAGGCGCATCAAGTACGAGCTAGCGGCAAGCTACCTTCA 1336
Db 1797 CCGGCGCGGATCCAGGCGGAGGCGCTCCAGGAGACCGCTATCGCGGAGCGGCTCTCG 1856
QY 1337 CCGGCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1386
Db 1857 TGAGTTCGCGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1906

RESULT 10

VCA429230

LOCUS

VC429230

DEFINITION

Volvox carteri f. nagariensis mRNA for perophorin-d21 protein.

ACCESSION

AJ429230

VERSION

AJ429230.1

KEYWORDS

perophorin-d21 protein.

SOURCE

Volvox carteri f. nagariensis.

ORGANISM

Volvox carteri f. nagariensis.

REFERENCE

1

AUTHORS

Ender, P., Godl, K., Wenzl, S. and Sumper, M.

1

Query Match 9.6%; Score 190; DB 6; Length 2539;
Best Local Similarity 54.4%; Pred. No. 5.7e-15;
Matches 582; Conservative 0; Mismatches 440; Indels 48; Gaps 8:

QY 332 TGATAGACACGGCTTACTTCGGGTAAACACAGGGTCTCGGGCATAAAGTTTCATACAGGAGG 391
DB 870 TGTGGAGCGCGTACACAGCGGACCGGACCGCTCCGGCATCGCCACGCTCGCTGG 929
QY 392 ATTACAGGTTTACAGTTGACGAGCGGACCTTCGCTTCCAGATAGGCGCGGATACCGTCT 451
DB 930 ACGGGTCCGAGGCGCGCTTCGACACGCTCGTTCGGGACAGATCGCGCGCCCAAGCGGT 989
QY 452 GSAATCTCCCTCGGCTACGACGGAAGCGGTGCTGCTTCCGATCGCTGATAGGATAG 511
DB 990 GG---TCCGCGGCTACGAGCGGCAAGGGCGTGAAGATCGCGCTCTGGACACCGGTGTCG 1046
QY 512 ACGGACACCGCGATCTGAGGGCAAGGTTCATAGGCTGTGACGACCGCGTCAAGGCA 571
DB 1047 ACGAGGACATCCGACCTGAGGGCGGGTGA---CCGGTCCAAGAACTTCACCGCG 1103
QY 572 GGTGACCCCTACGATGACGAGGACGGAACCCAGCTTCCGGTATCGTTGCGGAA 631
DB 1104 CCGCGGCGCGGACAGAGTGGCGCACCGCACCGCTCGCTCGATCGCGGGGCA 1163
QY 632 CCGGCAC---GTTAACTCCAGTACATAGGCTCGCGCGCGGCGGAGCTGTCGCG 688
DB 1164 CCGGCGCGGACGTCACAGGGCAAGTACAGGGCGCTCGCACCGCGCGGCGATCCTCAAG 1223
QY 689 TCAAGTCTCGGTCGCGGAGCTTCGGAAGCGCTTCACATCATCGCGGCTGTTGACT 748
DB 1224 GCAAGTCTCGACGACTCGGTTTCGGGACGACTCGCGCATCTCCGCGCGATGAGT 1283
QY 749 GGGTGTCCAGACAGGACAGTACGAGTAAAGGTCATCACTCTCCCTCGCTCT 808
DB 1284 GGS-----CGGCGCGGACGCGGCGGCGGCTGTCACCATGAGCTTGGCGGCA 1331
QY 809 CCGAGGCTCGGAGGAAACGACTCCCTAGTCAAGCGCGTCAACACGCTGGGACGCG 868
DB 1332 TGGACACACCGGAGGACCGACCGCTGGAGCGCGGCTCGAACAGCTCGCGCGAGAGG 1391
QY 869 GTATAGTGTCTGCTCGCGCGCGGACAGCGCGGCGGACACCTACACGCTCGCTCAC 928
DB 1392 GGTCTCTGTCGCTCGCGCGCGGCGGACGAGGCGCG-----GAGTGTGCTGCTGCG 1445
QY 929 CCGCGCGCGGAGGAGGTCATACCGTGGTGTGAGTGTGACAGCAAGACATCGGCA 988
DB 1446 CCGGAGCGGAGCGCGCGCTCACCGTCCGCGCGCTCGACGACAGCAAGGACGCTCGCG 1505
QY 989 GCTTCTCCAGGAGGACCGAC---CGCGGACGGAAGCTCAAGCGGAGTCTGTCGCGC 1045
DB 1506 ACTTCTCCAGCGCGCGCGCTCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCTC 1565
QY 1046 CCGGCGTGTGACATAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1100
DB 1565 CCGGCGTGTGACATAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1625
QY 1101 ----AAAGGACTACTACCAAGGCTCTGGAACGAGCATGGCGGCGGCGGCGGCGGCGG 1156
DB 1626 AGGAGCGCGCGGCTACATGACCATCTCGGACGCTGATGGCGGCGGCGGCGGCGGCGG 1685
QY 1157 GGTGTGGCGGCTCATCTCCAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1216
DB 1686 CCGCGCGCGGCGGCTCTGAGGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1745
QY 1217 CCGCTCATCAGACCGCGGAGATAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1276
DB 1746 CGCTCAGCGGCTCCACCAAGGCGGCG-----AAGTACACCGGCTTCGAGCGGCTT 1796
QY 1277 CCGGTAGGCTCAGAGGCTCAGAGGCTCAGTACGAGGCTACGAGGCTACGCGGCTCAG 1336
DB 1797 CCGGCGGCTCAGAGGCTCAGAGGCTCAGTACGAGGCTACGAGGCTCAGCGGCTCAG 1856

QY 1337 CGGCTCGCTCGCGACAGGAGCGCCACCCACACCTTCGAGCTGACG 1386
DB 1857 TGAGCTCGGCTCGCGACCTGCGCGCACCGACGAGCGCGGTGAC 1906

RESULT 12
D83672 3348 bp DNA linear BCT 06-FEB-1999
LOCUS Streptomyces albobogriiseolus DNA for subtilisin-like protease,
complete cds.
DEFINITION D83672
ACCESSION D83672
VERSION D83672.1 GI:1694626
KEYWORDS subtilisin-like protease.
SOURCE Streptomyces albobogriiseolus (strain:S-3253) DNA, clone:SAM-P45.
ORGANISM Streptomyces albobogriiseolus
Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (sites)
AUTHORS Suzuki,M., Taguchi,S., Yamada,S., Kojima,S., Miura,K.I. and
Homose,H.
TITLE A novel member of the subtilisin-like protease family from
Streptomyces albobogriiseolus
J. Bacteriol. 179 (2), 430-438 (1997)
JOURNAL 97144528
MEDLINE 2 (bases 1 to 3348)
REFERENCE Taguchi,S.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 3348)
AUTHORS Taguchi,S.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1996) Selichi Taguchi, Science University of
Tokyo, Dept. of Biological Science and Technology, 2841 Yamazaki,
Noda, Chiba 278-8510, Japan (E-mail:taguchi@rs.noda.sut.ac.jp,
Tel:0471-24-1501 (ex.4428), Fax:0471-25-1841)
COMMENT Sequence updated (07-Nov-1996) by : Selichi Taguchi.
FEATURES
Location/Qualifiers
1. 3348
/organism="Streptomyces albobogriiseolus"
/strain="S-3253"
/db_xref="taxon:1887"
/clone="SAM-P45"
2. 7
/db_xref="GI:1694627"
14. 3322
/note="molecular masses observed on SDS/PAGE: 45kDa and
40kDa"
/codon_start=1
/transl_table=11
/product="subtilisin-like protease"
/protein_id="BAAL2040.1"
/db_xref="GI:1694627"
/strand="+"
/translation="NRGDRNGDRVALAAGMAGPASPANGENSTAGSSATALKGHI
RYLTITGDFVALDAGKRYVLEPAGREHPIVDIRSDGHTLTPADARLVAAGKLD
QRUFDTLEKNAATRAHGLAVYVGTGCAAKAAKADVRDGTVRRTLVSLNADAVQ
TPENAGAEIWEQVTDGRTASGARVWLDGVRKASLDTSVGIGTPKAMEAGYDQGV
KIAVLDTGDVATHDPLKQVTSKNFTSPTTGDVYVGHGTHVSIAGTGAQSKTVK
GVAPGAKILNGKVLDDGDDSGILAGMEWAAQADIVNNKSLGCHDPTDPLEAA
VVKLSAEKILFAIAAGNEGPGSIGSPGSADSLTYGVDDKDLADPSSTGPRLGQ
AVKPDLTACPDVITAAAGKNDIAKEVEKPGMYTISGTSMATPHVAGAAALLKQOH
PEKVAELAGALTASTKDGKVTPEEGSGRVOVDKAITOTVIAEVPVSLPGVQMPHA
DDKPVTKLTYNLGCTEDYTLKLTATGKGAAPNGEFTLGASTLTPANGTASVD
VTADTLGAVNCTTSAYVATGAGSYRTAAAREVESYNTLYLDRSGKATANT
MAYLSGTGLGKDRSTAPTEADGAVSYRVPNGGTVLDASVLVGADEPTWGRADLAP
KLDVTRNTTVDKAKAPKPVTPGKAQAQYTAISGGKVKLATGTATYDTSVYGMLENYSG
FRSHLGPQITNTGLSQMHPFNGAKAQYTAISGGKVKLATGTATYDTSVYGMLENYSG
QVNGGAAGKKGKAVTAFGLPGSSGASGSEKLPSTRTLYLSTVNGVMDLDFPQ
LGVDNEGHPIDAVTITGVGKTYKAGTKVETNVAFGPRLTYSYGVGFRDGNISYV
VIFLADGKHAGSESSAVTLYRNGKAVGNNDLPFCBEGFTVPVSDAAAYRLTTS
VKRSKVAASATRIDASPTFRSKTSEKOLPVSSARFAATGLDSKVAAGKATPV
VPCAAQGNLKLAVYVSYNGGKTHTTKITVKNPAKKAISFRAKITDKKG
NASLTIHNAITGK.
594 a 1234 c 1133 g 387 t


```

/ gene="Pdx2"
/ note="TFE0822"
/ complement(2778. .3347)
/ gene="Pdx2"
/ note="Best Blastp hit = gi|17447528|pir|I172371|
hypothetical protein TM0472 - Thermotoga maritima (strain
MSB8) gi|14980982|gb|AA035557.1|AE001725_2 (AE001725)
amidotransferase, putative [Thermotoga maritima], score
158, E-value 6.00E-38"
/ codon_start=1
/ trans_table=11
/ product="predicted glutamine amidotransferase involved in
pyridoxine biosynthesis"
/ protein_id="AA024079.1"
/ db_xref="GI:20515816"
/ translation="MRVGLGVGCGSKHMDKLKIKGIEAVEAKDKDULLLDLALII
PGESTAGKLVDFGLKDALIKNERKVPFWGTGAGMLAKHIVDNRHRLGIMDI
TVRRAYGSGNSFKRLIIPAISENEIEAFIRAPYIESVGDGVRILAKYEGKIVAA
QOCNLTAFHPDLTDLSPRYFLRLNS"
/ complement(2781. .3335)
/ gene="Pdx2"
/ note="Best Blastp hit = gi|1176057|sp|P45293|YGA7_HAEIN
HYPOTHEICAL PROTEIN H11647 gi|1074877|pir|I164173
hypothetical protein H11647 - Haemophilus influenzae
(strain Rd KW20) gi|1574495|gb|AAC23294.1| (U32837)
conserved hypothetical protein [Haemophilus influenzae
Rd], score 392, E-value 1.00E-108"
/ codon_start=1
/ trans_table=11
/ product="pyridoxine biosynthesis enzyme"
/ protein_id="AA024080.1"
/ db_xref="GI:20515817"
/ translation="MNSRYELNKLKMLKGGVINDVTPSEALIAKACAVAMALE
RVPDILARGVARNSDPKIIEKAAVSPYMAKVRGHEVMAQLLEAGLIDFIDES
EVUTADENYHINWDFVPYFGKAGNLGALRIAGSRMGKAGGAGNVEYVR
HWRMNAELKRLITLGEEDMAAKELQAPYDLVYVAHQGLPVPVAGGATMPAD
AALMWLGADGVFGSGIFSONPEKMAAIAKAVTYTDKPEILAEVSEGLGEATMDI
DRLQLEKDLVATRGW"
/ complement(3356. .3360)
/ gene="Pdx2"
/ complement(3598. .4218)
/ gene="Snz1"
/ note="Best Blastp hit = gi|17475581|pir|I169587|
intracellular alkaline serine proteinase aprx - Bacillus
subtilis gi|12634110|emb|CAB13610.1| (F99113) alkaline
serine protease [Bacillus subtilis], score 210, E-value
4.00E-53"
/ codon_start=1
/ trans_table=11
/ product="Subtilisin-like serine proteases"
/ protein_id="AA024081.1"
/ db_xref="GI:20515818"
/ translation="MKKQLAKILLSLALIIISLISNELVQAQPNQINIPSPSKI
YPSLPQISQNSKNKIFDDEQLRLNKPDSSEFFVITFNKPYSDADITAKNIG
KPKIKRYKIIIPSIANAUTKQINVLKSLKIEIVQIEYDEPVATLDATKWFQTKAR
SDQVYTGKNTIATIDTGDGNHVDLSGKIIIGMKDFINNKTPYDDNGHGTWASIA
AGTGAGNSFYKGVAFDALLVGIKVLDSGSGMSHTVYAGIDMAVONKDVGIKVLNS

```

```

LGTSTSDGTDSTSLAVNRVDSGVVVAAGNSGPAKYITGSPGAEEKAITVAAMAD
VGLGNFLASFSGRGTADGRIKPDIAAPGNITAAKANSVNGYVYVYSGTSNATPFA
GTVALKLNANPLTNDARKNIMSTAKSGPPSKNVDYAGRLDGYEARIRAGNFRGN
NDVPHVYISGYLPGSRYSOTWTFNATNTSYPIATLIIIPOMANYNPDFDIYLYDPS
GTLIKSQRTQREITILPSTCTGYVVKVSYRSGNYFFDLDSAGSGSPVKY"
/ complement(5356. .6192)
/ gene="AprE2"
/ note="Pfam match to entry Peptidase_S8, Subtilase family,
score 212.5, E-value 5.90E-60"
/ complement(6694. .6698)
/ gene="AprE2"
/ note="PdpF4"
/ gene="PdpF4"
/ note="TFE0825"
/ protein_id="AA024082.1"
/ db_xref="GI:20515819"
/ translation="MVISMKNDKQYVYVVEKSKNFLEDFVLSKGAATTHBIRIL
NLARSNYSAIFNLNVEEIKIIVSKPYVVEKIEIFQAFKRIKYEINPSELIINVS
BIKREYVAFGPNIIIGYNGSLQEQAAKAAVYVPPNGLITIIIGPSSGKTYLAES
MYKFRKGLKGSFEVLCADYVYHNPQLLSHLFCHAKGAYTCATFVERIGLVKANHG
ILFDEVRHLPPECQEMFLYLDKGMKRLGDTDEKRVKNYIAATTAETIDSLTRTF
IRRPVLIQIIPSDERPEVEKMEIIVILLREESKRLQVDIIISILLALLSLFKYEGN
IGELSKLIQIICSKARLANFOINRAFIKLSALLPSKYKELINTETDFEFLITPP
GKRSDIRYDIEFIACKYHELAKSGYEETIRDLIIVSNFLDDV"
/ 7455. .7814
/ gene="PspF4"
/ note="Pfam match to entry sigma54, Sigma-54 interaction
domain, score 77.4, E-value 1.50E-20"
/ complement(8414. .8657)
/ gene="TFE0826"
/ misc_feature
Query Match 9.18; Score 179; DB 1; Length 14376;
Best Local Similarity 55.6; Pred. No. 9.1e-14;
Matches 394; Conservative 0; Mismatches 300; Indels 15; Gaps 2;
QY 374 TAAAGTTCATACAGGAGGATTACAGGTTTCAGGTTGACGACCCACCTTCCTCCGTCCTCCAGA 433
DB 6281 TAAAGCAAAATTCAGTACCATGAGCTCTATATGCTACTTATAGATACCGCAACAAAGTGT 6222
QY 434 TAGGGCCGATACCGTCTGGAACCTCCGCTACGACGGAAGCGGTGCTGCTGCTGCCA 493
DB 6221 TTGGAAATCACAAAGACGCGGAGTGATTTTGGAGTCACAGGAAATAATATACAAATAGCA 6162
QY 494 TCGTCGATACGGGTATTCAGCGTGAACCCACCCGATCT---GAAGGCGCAAGGTGATAGGCT 550
DB 6161 TTATTACACAGGTATACAGCGAAATCAGCTTCACCTCTCAGGTGCAAAAATAATAGCAT 6102
QY 551 GGTACGACCGCTCAACGCGAGGTGCGACCCCTACGATGACGACGACGACGACGACGACG 610
DB 6101 GGAAGACTTTTCAACAAACAAACTACCATACGACGACGACGACGACGACGACGACG 6042
QY 611 TTCGGGTATCGTTCCGCGAACCSCGCGCGCTTAACCTCCAGTAGTAGGCTGCGCCCGCG 670
DB 6041 TAGCAAGTATTCGTCAGCTACAGGTGCTGGAACACAGTCTTTTACAAAGCGGCTGCTGCTG 5982

```


QY 1299 GCCATCAAGTACGACGACTACGCCAGCTCACCTTACCGGCTCGTCCGCGACAGGG 1358
 Db 20740 CGTACCCGACGAGCGGACGCCGACCTGGGCTTCTACTCTGCGCTGACGCGCGAGCA 20681
 QY 1359 AAGCGCCACCCACACCTTCAGCTCAGCGGGGCGACCTTGTGACGCGCACCTCTACT- 1417
 Db 20680 GCCGTACACGAGGCGCTGCTACACCACTCTCTGGACAGGAGCGCTGAGCTGAAGCT 20621
 QY 1418 -----GGACACGGGCTGAGGACGACATCGACCTTACCTTACGACCCCAAGCGAA 1469
 Db 20620 GTCCGTGCGGGGCGCCCGGAGGCGTCCGACGCTCGCGACACCTCCCTCACGTCGCC 20561
 QY 1470 CGAGCTTACCTCTCTACACCGCTTACTACGCTTACGCTTACGAGGCTGGGTACTACACCC 1529
 Db 20560 CGCGACGCGGACCGCTCGACCAAGCTCACCGGCGAGGCTTCAAGCGCGCTCGCGCA 20501
 QY 1530 GACCGCGGACCTGGAGGCTCAGGTCTGACGTACAGGCGCGCGGCGGACCTACCAAGT 1589
 Db 20500 CACGAGCGGCGAGCTGTGGGCGCGGAGCGGAGGCGGAGGCGGCTCGCGACACCGCTT 20441
 QY 1590 CGAGCTGTGACGCGGAGGCTCAGCGAGTCCGCGGCGGCGGACCGCAATCCAAACCC 1649
 Db 20440 CGCGCTGTGACGAGGCGGAGGCTACAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 20381
 QY 1650 CAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1709
 Db 20380 CGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 20321
 QY 1710 CTGGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1743
 Db 20320 GGCCACGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAC 20287

RESULT 15
 SC8A11/c 32704 bp DNA linear BCT 12 MAY 2002

LOCUS Streptomyces coelicolor cosmid 8A11.
 DEFINITION ALJ91041 AL645882
 ACCESSION ALJ91041.2 GI:20320858
 VERSION amino acid transport integral membran protein; branched amino acid
 KEYWORDS binding secreted protein; branched amino acid transport system
 ATP-binding protein; branched amino acid transport system permease;
 integral membrane protein; ion transport integral membrane protein;
 polysaccharide biosynthesis protein; regulator; regulatory protein;
 secreted amidase; secreted peptidase; secreted protein; signal
 factor; transcriptional regulator; transcriptional regulatory
 protein.

SOURCE Streptomyces coelicolor A3(2).
 ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 32704)
 Redenbach, M., Kleser, H.M., Denapate, D., Eichner, A., Cullum, J.,
 Kinashi, H., and Hopwood, D.A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 97000351
 PUBMED 8843436

REFERENCE 2 (bases 1 to 32704)
 Saunders, D.C. and Harris, D.
 JOURNAL Unpublished
 3 (bases 1 to 32704)
 Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
 Direct Submission
 Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK
 On May 9, 2002 this sequence version replaced gi:9716211.
 Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded

by the BBSRC and Beowulf Genomics details of S. coelicolor sequencing at the Sanger Centre are available on the world Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of sibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh.hgo.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 8A11.
 FEATURES
 source
 1. 32704
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid 8A11"
 2. 103
 /note="nominal overlap with Streptomyces coelicolor cosmid St8A11"
 misc_feature
 2043. 2240
 /note="Pfam match to entry PF00196 GrE, Bacterial regulatory proteins, luxR family, score 49.10, E-value 9.7e-11"
 misc_feature
 2094. 2177
 /note="PS00622 Bacterial regulatory proteins, luxR family signature"
 gene
 complement(2185..2454)
 /gene="SC8A11.02c"
 /note="SC07174"
 complement(2185..2454)
 /gene="SC8A11.02c"
 /note="SC8A11.02c", unknown, len: 89 aa"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical protein SC8A11.02c"
 /protein_id="CAC01574.1"
 /db_xref="GI:9716213"
 /db_xref="SPTREMBL:Q9FC08"
 /translation="MASAAATAGTAGALPVMGQAHADTRGLPSLMDYDVSYYVPENA YTVVDVQVRAEVAQLMGATEEETPRSRAGLRAPRRACRGW"
 2788. 3258
 /gene="SC8A11.03"
 /note="SCC7175"
 2788. 3258
 /gene="SC8A11.03"
 /note="SC8A11.03, possible regulator, len: 156 aa; similar to TR:O9S2F7 (EMBL:AL096852) Streptomyces coelicolor putative regulator SC819A.24, 142 aa; fasta scores: opt: 272 z-score: 353.0 E(): 3.4e-12; 36.9% identity in 130 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="putative regulator"

/protein_id="CAC01575.1"
 /db_xref="GI:9716214"
 /db_xref="sptrembl:O9FC07"
 /translation="MSDHHGVQAHSAQAALLPLSLCSQVCALEWEDYPAERFYD
 PPHLVLIRPEACAPVTHVHROLLHGLRTTSGLCGVQVADPTDRATWLVQ
 NAGHDIAFLPVPPELEWIDRTYLHVPAQTSSRLGTDAFLSKLDFEAPSR"
 complement(3298..7059)
 /gene="SC8A11.04c"
 /note="SC07176"
 complement(3298..7059)
 CDS
 /gene="SC8A11.04c"
 /note="SC8A11.04c"
 aa: similar to TR:Q9RL54 (EMBL:AL121596) Streptomyces
 coelicolor probable secreted peptidase SCF1A.10, 1245 aa;
 fasta scores: opt: 1827 z-score: 1848.3 E(): 0; 37.2%
 identity in 1300 aa overlap, to TR:P95684 (EMBL:D83672)
 Streptomyces albobogiseolus subtilisin-like protease, 1102
 aa; fasta scores: opt: 1365 z-score: 1380.6 E(): 0; 33.4%
 identity in 1268 aa overlap and to Streptomyces coelicolor
 probable secreted peptidase SC8A11.16c, 1239 aa; fasta
 scores: opt: 1835 z-score: 1467.4 E(): 0; 46.9% identity
 in 1279 aa overlap. Contains Pfam matches to entries
 PF00082 Peptidase_S8, Subtilase family and PF02225 PA, PA
 domain and matches to prosite entries PS00136 Serine
 proteases, subtilase family, aspartic acid active site,
 PS00137 Serine proteases, subtilase family, histidine
 active site and PS00138 Serine proteases, subtilase
 family, serine active site. Also contains possible
 N-terminal region signal peptide sequence.
 /codon_start=1
 /transl_table=11
 /product="putative secreted peptidase"
 /protein_id="CAC01576.1"
 /db_xref="GI:9716215"
 /db_xref="sptrembl:O9FC06"
 /translation="MPLPNSRGRLTVGIALLVLTATAPASLPCTVPPAPTAPA
 GTRVLITIDGRTVPTVGGAEVLVSDPHRSGGHVMTVGSDFYVYVPAAPYL
 GGADELRFVTELEDGDDADDELPLVYTDTAARSLGARTPEGARTRALS
 IGRGAIISAEHRAADFWTSITGCDAAAGCSAARSATSGRLAGGIKVLWGDKVAT
 LSDTQAGTAPVWNGTGEVGVAVLDTGDAGHPDFAGRIATAAFSPVDQVYDR
 NGHTHVASTVAGTGAAGGVGVAPGASLHIGVLDNSGSDSVYLAGHEWAVDR
 OIAKIVNSLSDSPDGTGDTPLSEAVNLSAFTGALFYVAGNSGPEAYTVGPAPADA
 ALTGAVNGPKGVQDLADESSRPVGNVAKPDLTAPGVGLAARSRYAPDEGAY
 QLSGTSHATPHYAGAAALIAEHPONTGRLKALYGTGTAGTORTSPFDGSGRDV
 AAARSTLLSDGAPAHYPTFGQTVRDVYTNNSGPAPVALDLSALPELPEGLF
 TUSAQVTPVAPGATSVGVITHLDAEDNGAYATRLVAGADGAVLATPVPVMEGR
 RATLALTARDHDKPLSGTVILKIDVERNTAPKVSYVDASGRDLRLSPSYVMKNSA
 VPGVGTHTLGFAMFTAPVYLDADRTVAPDADDLRKAAYPRATANORLIDQYRG
 NTGLPFPMDSYVAEYWRVDSLVTPTPEVKTGSYTFATRMROIOPPLTFAGSOTFDD
 VIVQSRSPQLPESTRAYRVAGDGSATEFGEAEVDRVAVVRESDTVPATQDAAAE
 KAPRYQLLNDYGVKFPDPAHLPEAPLPVASLGTDDSAHLARFPGAGTTLVRS
 HGVRYAYDLVYRHHDGAVPRDPSYRPARGELARVDDTFRDTSQGRAVEYHDIISLGG
 PLGVPTQVRAQELTISWTADDDVWVSFASRPDLOGRVARSIEPTSTRTIFAP
 IQHPRLLSDNTSGGPFPRAGDNISVTANGSDGGHGVYVWADGTIRISLYOGGE
 LGEDVNERIVNVGSLSPGPKPYRLVLEGRNLDPDPTSTRTVNDFTSATDPTRL
 TPLVLVQLDYAVAVDLSGRAHRRTELTVTASHLEGAAGAGARTATVEVSDDGATW
 RTALRSAGWTARLDAPGARVARSILRTAKDTEGNGVGTLRFAEGLR"
 complement(4375..4614)
 /gene="SC8A11.04c"
 /note="Pfam match to entry PF02225 PA, PA domain, score
 19.00, E-value 0.058"
 complement(5521..6417)
 /gene="SC8A11.04c"
 /note="Pfam match to entry PF00082 Peptidase_S8, Subtilase
 family, score 199.50, E-value 5e-56"
 complement(5665..5697)
 /gene="SC8A11.04c"
 /note="PS00138 Serine proteases, subtilase family, serine
 active site"
 complement(6193..6225)
 /gene="SC8A11.04c"
 /note="PS00137 Serine proteases, subtilase family,
 histidine active site"

misc_feature complement(6298..6333)
 /gene="SC8A11.04c"
 /note="PS00136 Serine proteases, subtilase family,
 aspartic acid active site"
 RBS complement(7064..7069)
 RBS 7287..7290
 gene 7295..7339
 /gene="SC8A11.05"
 /note="SC07177"
 CDS 7295..7939
 /gene="SC8A11.05"
 /note="SC8A11.05, unknown, len: 214 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SC8A11.05"
 /protein_id="CAC01577.1"
 Query Match 7.5% Score 149; DB 1; Length 32104;
 Best Local Similarity 50.5% Pred. No. 4.3e-10;
 Matches 527; Conservative 0; Mismatches 475; Indels 33; Gaps 6:
 QY 381 CATACAGAGGATTACAAAGTTTCAGGTTGACGACGCCACTTCCCTCTCTCCAGATAGGGC 440
 DB 21772 CATCTGGCTCCACCCCGCGGTGAGGCCGACCTGSCCGACAGCACGATCGCGC 21713
 QY 441 CGATACCGCTTGGAACTCCCTCGGCTACGACGGAAGCGGTGTGGTGTGCCATCTCGA 500
 DB 21712 CCGGAGGCGCTGG---GCCCGGGCAACACCGGGCGAGGGGTGGAGGTCGCCGTGCTGA 21656
 QY 501 TAGCGGTATAGACGCGAACCCACCCGATCTGAAGCGCAAGCTCATAGGCTGCTACACGCG 560
 DB 21655 CACCGCGCTCCAGCGGGGACCGCGACCTCGCGCGCGGACAGGCTT 21596
 QY 561 CGTCAACGGCAGGTGACCCCTACGATGCCAGGACACGGAACCCACGTTTGGGGTAT 620
 DB 21595 CGTACCGGACGCA---GAACACCGACCGCGGACCGGACCGGACCCACGTTGGCTCCAC 21539
 QY 621 CGTTGCCGGAACCGCAGCGGTAACTCCG---AGTACATAGCGGTCCGCCCGCGCGCA 677
 DB 21538 CATCGCGGCAACCGGTGGGCTCCCGCGCAAGGAGAGGGGTGGCGCGCGGACG 21479
 QY 678 GCTCGTGGCGCTCAAGGTTCTCGTGGCGACCGTTTCGGGAGCGCTTCCACCATCTCC 737
 DB 21478 GCTGAGTATCGGAGAGGTGTGGCAACATCCCGGCGCGGCGAGATCTCTTGGAGCGTGGC 21419
 QY 738 GGGTGTGACTGGGTCTCTCCGAACAGGACAGTACGCGGATAAGGTCTATCAACCTCTC 797
 DB 21418 CGCCATGGAGTGGCGG-----CTGTCAACCGCCACCGCAAGATCGTCAACATGAG 21368
 QY 798 CTTGGCTCTCCCGAGAGTCCGACGGAACCCACTCTCCCTCAGTCAGCGCGCTCAACAACGC 857
 DB 21367 CTTGGCTCTCCCGTACGAGTCCGACGCGACCGATCCGATGAGCGCGCGCTTGGACGGCT 21308
 QY 858 CTGGGACGCGCGGTATAGTATCTGCTGCCCGCGCGCAACAGCGCGCGGCGCAACCTTACAC 917
 DB 21307 CAGCGCCAGACCGCGCGGTGTGTCTGCTCGCGCGCGCAACAGCGCGCGGCGCGCGCAG 21248
 QY 918 CGTGCGCTCACCGCGCGCGCGGAGCAAGGTCTATACCGTCTGCTCAGTTCACAGCAACG 977
 DB 21247 CATCGCGCGCGCTGTGTATGCCACTCTCCCGGTGTACCTTGGCGCGGTTCACGCGCACCA 21188
 QY 978 CAACATTCGCGAGCTTCTCCACGAGGAGCGGACCGCGGACGGAAGGTCTCAAGCGCGAAGT 1037
 DB 21187 CAGCGTCTCGCGCTTCTCCACGAGCGCGCGCGCGGTTCGACGCGCGGTGAAGCGCGAGAT 21128
 QY 1038 CGTGGCGCGCGCGGTGTACCTATAGCGCGCGCGCGGACGGAACAGCATGGGCAACCC 1097
 DB 21127 CACCGCGCGCGGTGGCATCTCGCGCGCAACCTCTCTTCTCGCGCGCGCGCGG----- 21073
 QY 1098 GATAAAGCATCTACTACACGAGGCGCTCTGGAACACGAGCATGGCGACCGCGCGCGCGG 1157
 DB 21072 ---AACGTCGCTACGAGAGCGCTGAGCGCGCACCTCGATGCGCCACCGCGCGCGCGCGG 21017

QY 1158 CGTTGGCGCTATCTTCAGCCGCCAGCTGAGCCCGGACAGGTGAAGACCGG 1217
DB 21016 CGCCCGCGCTGCTGCCCGCGCAGCTCCCGACCTCAGCGGCGCCCTCAAGGCGT 20957
QY 1218 COTCATCGAGACCGCGACATAGTCCGCCCAAGGAGATAGCGGACATCGCCTAGCGTGG 1277
DB 20955 COTCCCGAGAGTTCCGA-----CGGCACITCCGGGTACGACGCTTCCAGGCGGCGAG 20903
QY 1278 GGGTAGGTGAAGCTTACAGGCCCATCAAGTACGACGACTAGCCCAAGCTCACCTTCAC 1337
DB 20902 CGGCCGGGTGACCTGACCGCCCGGCTGCGCGCGGGGTGACGCTCGGCGACCGCCTA 20843
QY 1338 CGGCTCCGTCGCCGACAAAGGGAAGCCACACACCTTCGAGCTCAGCGGCGCCACCTT 1397
DB 20842 CGCCCGGGCAGCTCACCGGCGCGCTGCGGCGCTGCTGACGTACACGACACACCGG 20783
QY 1398 CGTGACCGCCACCT 1412
DB 20782 CGCAGCGTCACT 20768

Search completed: November 2, 2002, 10:23:32
Job time : 3523.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2002, 08:16:03 : Search time 51.5 seconds
(without alignments)
11772.815 Million cell updates/sec

Title: US-09-841-553-2

Perfect score: 1977

Sequence: 1 ATGAAGAGGTAGGTGCTGT.....AGGCGTGTCTACTACGGG 1977

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents: NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1977	100.0	1977	4	US-08-894-818B-2
2	1360.8	68.8	1977	4	US-09-445-472-11
3	1165	58.9	1859	4	US-08-894-818B-6
4	833.2	42.1	898	1	US-08-894-818B-15
5	806.8	40.8	1464	4	US-08-750-532-7
6	750.2	37.9	1962	4	US-08-894-818B-17
7	750.2	37.9	1962	4	US-08-894-818B-34
8	692.8	35.0	1962	4	US-09-445-472-15
9	560.4	28.3	1236	4	US-08-894-818B-4
10	532.8	26.9	564	1	US-09-445-472-2
11	532.8	26.9	564	1	US-08-750-532-11
12	532.8	26.9	564	4	US-08-894-818B-14
13	191.6	9.7	2809	3	US-09-000-016-1
14	191.6	9.7	2809	4	US-09-514-340-1
15	190	9.6	2539	3	US-09-000-016-3
16	190	9.6	2539	4	US-09-514-340-3
17	145.4	7.4	507	4	US-08-894-818B-26
18	125.4	6.3	1140	3	US-09-023-173-4
19	124	6.3	145	1	US-08-750-532-10
20	124	6.3	145	4	US-08-894-818B-13
21	113.2	5.7	4403765	4	US-09-103-840A-2
22	107.6	5.4	5163	3	US-08-700-651-1
23	107.6	5.4	5163	3	US-08-928-361B-4
24	107.6	5.4	5318	3	US-08-700-651-2
25	107.6	5.4	5318	3	US-08-928-361B-3
26	104.2	5.3	2712	3	US-09-025-691-4
27	97.4	4.9	5511	3	US-08-928-361B-2

28	97.4	4.9	7334	3	US-08-928-361B-1
29	86.8	4.4	2064	1	US-08-343-428-1
30	84.8	4.3	350	4	US-09-197-649-7
31	83.8	4.2	1209	1	US-08-314-309A-5
32	83.8	4.2	1513	1	US-08-314-309A-2
33	83.8	4.2	1546	1	US-08-314-309A-3
34	80.6	4.1	2576	1	US-08-471-033-35
35	80.6	4.1	2576	2	US-08-471-044-35
36	80.6	4.1	2576	2	US-08-463-483A-35
37	80.6	4.1	2576	2	US-08-471-046A-35
38	80.6	4.1	2576	2	US-08-470-366B-35
39	80.6	4.1	2576	2	US-08-463-334-35
40	80.6	4.1	2576	3	US-09-300-529-35
41	80.6	4.1	2655	1	US-08-471-033-26
42	80.6	4.1	2655	2	US-08-471-044-26
43	80.6	4.1	2655	2	US-08-463-483A-26
44	80.6	4.1	2655	2	US-08-471-046A-26
45	80.6	4.1	2655	3	US-08-470-566B-26

ALIGNMENTS

RESULT 1
US-08-894-818B-2
Sequence 2, Application US/08894818B
Patent No. 5261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Iiyaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Keisuke
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshi
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
Zip: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELEPHONE: (202) 638-5197
TELEFAX: (202) 737-3533
INFORMATION FOR SEQ ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA

Does not claim
seq. D.2.
seq. D.2. seq. D.3
or claim for a polypeptide
of the amino acid
this is the end of the
sequence

US-08-894-818B-2

Query Match		100.0%; Score 1977; DB 4; Length 1977;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 1977; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGAAGAGTTAGTGTCTGTGGTGTGGCACTGTGTCTGGTCTTCTGCGCGGACG	60
DB	1	ATGAAGAGTTAGTGTCTGTGGTGTGGCACTGTGTCTGGTCTTCTGCGCGGACG	60
QY	61	GCCCTTGGCGCACCCGTAACCGGTTGTCAAGAACACCGGTTTCAGCAGAGAACTAC	120
DB	61	GCCCTTGGCGCACCCGTAACCGGTTGTCAAGAACACCGGTTTCAGCAGAGAACTAC	120
QY	121	GCAGTCTGACCCGCGGACTGTTCAGAAAGTCCAGAGGATGAACAGCAGAAAGT	180
DB	121	GCAGTCTGACCCGCGGACTGTTCAGAAAGTCCAGAGGATGAACAGCAGAAAGT	180
QY	181	GACACCGTCAATATGTTCCGGAGCTACGGAGACAGGAGACAGGCGGTTAAGTACTGAGG	240
DB	181	GACACCGTCAATATGTTCCGGAGCTACGGAGACAGGAGACAGGCGGTTAAGTACTGAGG	240
QY	241	CTCATGGCGCCAGGTCAAGTACTCTTACAGATAATCCCTGTCTGCGGTTAAATA	300
DB	241	CTCATGGCGCCAGGTCAAGTACTCTTACAGATAATCCCTGTCTGCGGTTAAATA	300
QY	301	AAGGCCAGGACCTTCTGCTGATCGCGGCGATGATAGACAGCGGTTACTTCGGTACACA	360
DB	301	AAGGCCAGGACCTTCTGCTGATCGCGGCGATGATAGACAGCGGTTACTTCGGTACACA	360
QY	361	AGGCTCTCGGCATAAAGTTTCATACAGAGGATTACAAAGTTACAGGTTACAGACGCCACT	420
DB	361	AGGCTCTCGGCATAAAGTTTCATACAGAGGATTACAAAGTTACAGGTTACAGACGCCACT	420
QY	421	TCCGTCTCCAGATAGAGGCGGATACCGTCTGGAACCTCCCTGCTACGACGAGACGGT	480
DB	421	TCCGTCTCCAGATAGAGGCGGATACCGTCTGGAACCTCCCTGCTACGACGAGACGGT	480
QY	481	GTGGTGTTCGATCGATAGCGGTATAGACGCGAACCCCGCTCTGAAGGCGAAG	540
DB	481	GTGGTGTTCGATCGATAGCGGTATAGACGCGAACCCCGCTCTGAAGGCGAAG	540
QY	541	GTCATAGGCTGTAGCGCGGCTCAACGCGAGTCCAGCCCTACGATGACCAAGGACAC	600
DB	541	GTCATAGGCTGTAGCGCGGCTCAACGCGAGTCCAGCCCTACGATGACCAAGGACAC	600
QY	601	GGAAACCAAGTTGCGGGTATCGTTGCGGGAACCGGACGGTTAACTCCCAATACATAGGC	660
DB	601	GGAAACCAAGTTGCGGGTATCGTTGCGGGAACCGGACGGTTAACTCCCAATACATAGGC	660
QY	661	GTGCGCCCGGCGGAGCTGCTGCGGCTCAAGGTTCTCGTGCGGACGGTTTCGGGAAGC	720
DB	661	GTGCGCCCGGCGGAGCTGCTGCGGCTCAAGGTTCTCGTGCGGACGGTTTCGGGAAGC	720
QY	721	GTCTCCACCATCATCGCGGGTGTGACTCGGTCGAGAACAGGACAAAGTACGGGATA	780
DB	721	GTCTCCACCATCATCGCGGGTGTGACTCGGTCGAGAACAGGACAAAGTACGGGATA	780
QY	781	AGGCTCATCAACTCTCCCTCGGCTCTCCAGAGCTCCGAGGACCGACTCCCTCACT	840
DB	781	AGGCTCATCAACTCTCCCTCGGCTCTCCAGAGCTCCGAGGACCGACTCCCTCACT	840
QY	841	CAGGCGCTCAACAGCGCTGGAGCGCGGTATAGTCTCGTCTCGCGCGCGGCAACAGC	900
DB	841	CAGGCGCTCAACAGCGCTGGAGCGCGGTATAGTCTCGTCTCGCGCGCGGCAACAGC	900
QY	901	GGGCCCAACACCTACACCGTCTGCTCACCGCGCGCGGAGCAAGGTCATAACCGTCT	960
DB	901	GGGCCCAACACCTACACCGTCTGCTCACCGCGCGCGGAGCAAGGTCATAACCGTCT	960
QY	961	GCAGTTGACAGCAAGCAACATCGGCGGCTTCTCCAGAGGAGGACCGCGGAGCA	1020
DB	961	GCAGTTGACAGCAAGCAACATCGGCGGCTTCTCCAGAGGAGGACCGCGGAGCA	1020

QY	1021	AGGCTCAAGCGGGAAGTGTCTGCCCCCGGGCTTGACATCATAGCCCGCGCGGACGGGA	1080
DB	1021	AGGCTCAAGCGGGAAGTGTCTGCCCCCGGGCTTGACATCATAGCCCGCGCGGACGGGA	1080
QY	1081	ACCAGCATGGGACACCCCGATTAACGACTACTACCAAGGCTCTTGAAACACGACATGGCC	1140
DB	1081	ACCAGCATGGGACACCCCGATTAACGACTACTACCAAGGCTCTTGAAACACGACATGGCC	1140
QY	1141	ACCCGCGAGTTTCGGCGGTTGGCGGCTCATCTCCAGGCCCCACCGAGCTGGACCCG	1200
DB	1141	ACCCGCGAGTTTCGGCGGTTGGCGGCTCATCTCCAGGCCCCACCGAGCTGGACCCG	1200
QY	1201	GACAAAGTCAAGACACCGGCTTCATCGAGACCGCGGACATAGTCCGCCCAAGGAGATAGG	1260
DB	1201	GACAAAGTCAAGACACCGGCTTCATCGAGACCGCGGACATAGTCCGCCCAAGGAGATAGG	1260
QY	1261	GACATCGCTACGGTGGGCTAGGTTGAGCTTCAAGGCGCTCAAGTACGACGACTAC	1320
DB	1261	GACATCGCTACGGTGGGCTAGGTTGAGCTTCAAGGCGCTCAAGTACGACGACTAC	1320
QY	1321	GCCAAAGTCACTTCACCTTCGCTCCGCGGACAAAGGAAGCGCCACCCACACTTTCGAC	1380
DB	1321	GCCAAAGTCACTTCACCTTCGCTCCGCGGACAAAGGAAGCGCCACCCACACTTTCGAC	1380
QY	1381	GTCAGCGGCGCACCTTCGTCGACCGCACCTTACTGGGACAGCGGCTCGAGCGACATC	1440
DB	1381	GTCAGCGGCGCACCTTCGTCGACCGCACCTTACTGGGACAGCGGCTCGAGCGACATC	1440
QY	1441	GACCTCTACCTCTACGACCCACCGGGAACAGAGTTGACTACTCTACACCGCTTACTAC	1500
DB	1441	GACCTCTACCTCTACGACCCACCGGGAACAGAGTTGACTACTCTACACCGCTTACTAC	1500
QY	1501	GGCTTCGAGAGGTCGGGTATACAAACCGGACCGCGGACCTGGAGCGTCAAGAGTCTC	1560
DB	1501	GGCTTCGAGAGGTCGGGTATACAAACCGGACCGCGGACCTGGAGCGTCAAGAGTCTC	1560
QY	1561	AGCTACAGGCGGCGGCGGAACTACAGAGTTCGACGCTGTCAGGACGGGAGGCTCAGCG	1620
DB	1561	AGCTACAGGCGGCGGCGGAACTACAGAGTTCGACGCTGTCAGGACGGGAGGCTCAGCG	1620
QY	1621	TCCGGCGGCGGCAACCCGAAATCAACCCCAACCCGGAACCCCAACCCGCAACCC	1680
DB	1621	TCCGGCGGCGGCAACCCGAAATCAACCCCAACCCGGAACCCCAACCCGCAACCC	1680
QY	1681	CAGACCTTCACCGGTCGCTTACGACTACTGGACACCGACACCTTCACCATGAAC	1740
DB	1681	CAGACCTTCACCGGTCGCTTACGACTACTGGACACCGACACCTTCACCATGAAC	1740
QY	1741	GTCAACAGCGGTGCCACCAAGATAACCGGTCGACTGACCTTCGATACCTTCTTCAACGAC	1800
DB	1741	GTCAACAGCGGTGCCACCAAGATAACCGGTCGACTGACCTTCGATACCTTCTTCAACGAC	1800
QY	1801	CTGACCTCTACCTCTACGACCCCAAGGCAACCTCTGTTGACAGGTCACGTCGAGCAAC	1860
DB	1801	CTGACCTCTACCTCTACGACCCCAAGGCAACCTCTGTTGACAGGTCACGTCGAGCAAC	1860
QY	1861	AGCTACGACGCTCGAGTACGCGCAACCCCGCGGGAACCTTGAGCTTCTCTGCTTAC	1920
DB	1861	AGCTACGACGCTCGAGTACGCGCAACCCCGCGGGAACCTTGAGCTTCTCTGCTTAC	1920
QY	1921	GCCCTACGACCTTCGGCTGGGCGGACTACAGCTCAAGGCGCTGCTCTACTACGGG	1977
DB	1921	GCCCTACGACCTTCGGCTGGGCGGACTACAGCTCAAGGCGCTGCTCTACTACGGG	1977

RESULT 2

US-09-445-478-11

Application US/09445472

Sequence 11

Patent No. 6358726

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, MIO

Seq ID NO. 11 is not
claimed. No Dbl pat.

Query Match	100.0%	Score 1977	DB 4	Length 1977
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 1977: Conservative	0:	Mismatches	0:	Indels
Gaps	0:			
Qy	1	ATGAAGAGTTAGTGTGCTGTGGTCTGGCACTGGTGTCTGTGGGTCTTCTGGCGGAAGC	50	
Db	1	ATGAAGAGTTAGTGTGCTGTGGTCTGGCACTGGTGTCTGTGGGTCTTCTGGCGGAAGC	50	
Qy	61	GCCTTTGGGGACCCGTTAAACCGGTTCTCAGGAACAACCCGGTTTCACGAGAAGACTAC	120	
Db	61	GCCTTTGGGGACCCGTTAAACCGGTTCTCAGGAACAACCCGGTTTCACGAGAAGACTAC	120	
Qy	121	GGACTGCTACCCCGGGAGCTGTTCAAGAAAGCTCCAGAGGATGAATCGAACCAGGAAGTG	180	
Db	121	GGACTGCTACCCCGGGAGCTGTTCAAGAAAGCTCCAGAGGATGAATCGAACCAGGAAGTG	180	
Qy	181	GACACGTCATATGTTTCGGGAGCTACGGAGACAGGAGCAGGCGGTTAAGGTACTGAGG	240	
Db	181	GACACGTCATATGTTTCGGGAGCTACGGAGACAGGAGCAGGCGGTTAAGGTACTGAGG	240	
Qy	241	CTCATGGGGCCCAAGTCAAGTACTCTACAGATATTCCTGCTGTGGGTTTAAATA	300	
Db	241	CTCATGGGGCCCAAGTCAAGTACTCTACAGATATTCCTGCTGTGGGTTTAAATA	300	
Qy	301	AAGGCCAGGGACCTTCTGCTGATCGCGGGCATGATACACACGGGTTACTTCGGTAAACACA	360	
Db	301	AAGGCCAGGGACCTTCTGCTGATCGCGGGCATGATACACACGGGTTACTTCGGTAAACACA	360	
Qy	361	AGGCTCTGGGCATTAAGTTCATACAGAGGATACAAAGTTTCAGGTTGACGAGCCACT	420	
Db	361	AGGCTCTGGGCATTAAGTTCATACAGAGGATACAAAGTTTCAGGTTGACGAGCCACT	420	
Qy	421	TCGCTCTCCAGATAGGGGCGCATACCTCTCGAACTCCCTCGGCTACGACGGAGCGGT	480	
Db	421	TCGCTCTCCAGATAGGGGCGCATACCTCTCGAACTCCCTCGGCTACGACGGAGCGGT	480	
Qy	481	GTGGTGGTTGCCATCTCGATACGGGTATAGACGGGAACACCCCGCATCTGAAGGGCAAG	540	
Db	481	GTGGTGGTTGCCATCTCGATACGGGTATAGACGGGAACACCCCGCATCTGAAGGGCAAG	540	
Qy	541	GTCTATAGGCTGTACGACGGCGTCAACGGCAGGTGACCCCGCTACGATGACCCAGGGACAC	600	
Db	541	GTCTATAGGCTGTACGACGGCGTCAACGGCAGGTGACCCCGCTACGATGACCCAGGGACAC	600	
Qy	601	GGAAACCCAGTTGCGGTTATCTGTTCGGGAACCGGACGGTTAACTCCAGTACATAGGC	660	
Db	601	GGAAACCCAGTTGCGGTTATCTGTTCGGGAACCGGACGGTTAACTCCAGTACATAGGC	660	
Qy	661	GTCCGCCCGGGGGAGCTGCTCGCGCTCAAGTTTCTCGGTGCCGACCGGTTTCGGGAAGC	720	
Db	661	GTCCGCCCGGGGGAGCTGCTCGCGCTCAAGTTTCTCGGTGCCGACCGGTTTCGGGAAGC	720	
Qy	721	GTCTCCACCATCATCGCGGGTGTGACTGGGTCTGCCAAGACAAAGGATACGGGATA	780	
Db	721	GTCTCCACCATCATCGCGGGTGTGACTGGGTCTGCCAAGACAAAGGATACGGGATA	780	

QY 1861 AGCTACGAGCAGCTGAGTAGTACGCCAACCCCGCCGGGACCTGGAGCTTCTCTGCTTAC 1920
DB 1861 AGCTACGAGCAGCTGAGTAGTACGCCAACCCCGCCGGGACCTGGAGCTTCTCTGCTTAC 1920
QY 1921 GCCTACGACACCTACGCTGGGCGGAGCTACCAAGCTCAAGGCGCTGCTCTACTACGGG 1977
DB 1921 GCCTACGACACCTACGCTGGGCGGAGCTACCAAGCTCAAGGCGCTGCTCTACTACGGG 1977

RESULT 3

US-08-894-818B-6

Sequence 6, Application US/08894818B

Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Kio

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MITTA, Masahori

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,818B

FILING DATE: 20-MAY-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03253

FILING DATE: 07-NOV-1996

APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TAKAKURA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1977 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

US-08-894-818B-6

Query Match 68.8%; Score 1350.8; DB 4; Length 1977;
Best Local Similarity 83.9%; Pred. No. 2.2e-263;
Matches 1573; Conservative 0; Mismatches 292; Indels 9; Gaps 3;

QY 110 AGAAGAACTACGGCTCTGACCCCGGGAGCTGTTCAGAAAGTCCAGAGATGAAGTCTG 169
DB 107 AGAAGAACTATGGTCTGCTACCCAGAGCTGTTCAGAAAGTCCAGAGATGAAGTCTG 166
QY 170 ACCAGAAAGTGGACCGCTCATATGTTGGGAGCTACGGAGACAGGACGGGCGGTGA 229
DB 167 ACAGGAAATCAGCAGTAAATGTTGTAATTTGAAACCATAGGAAAGAAATTCAGTAA 226

QY 230 AGGTACTGAGGCTCATGGGCTCCAGGTCAGGTCAAGTACTCTCTACAAGATATATCCCTGCTG 289
DB 227 GAGTCTTTGAGTTAATGGGTGCAAAAGTTAGGTATGCTGTACCATATATTATACCCCAATAG 286
QY 290 CGTTTAAATAAAGCCAGGCACTTCTGCTGATCCGGGGCATGATA---GACACCGGTT 346
DB 287 CTGCGGATCTTAAGGTTAGAGACTTACTAGTCATCTCAGGTTTAAAGGGGTTAAAGCTA 346
QY 347 ACTTCGGTAAACAAAGGCTCTCGGGCATAAAGTTCATACAGAGGATTAACAGGTTCAAG 406
DB 347 AGCTTTCAAGTGTAGGTTTATCCAGGAAGACTACAAAAGTTTACAGTTTTCAGCAGAAATAG 406
QY 407 TTGACGAGCCCACTTCGGTCTGTCAGATAGGGGCGGATACCGTCTGGAAGTCCCTCGGCT 466
DB 407 AAGGACTGGATGAGTCTCGAGTCAAGTTATGCAACTTACGTTTGGAACT---TGGGAT 463
QY 467 ACACCGAAGCGGTGCTGGTGGTCCCATCGTCGATACGGGTATAGACGGAACACCCCG 526
DB 464 ATCATGGTCTTGGAAATCAATAGGAATTAATGACACTGGAATTGACGCTTCTCATCCAG 523
QY 527 ATCTGAAGSGCAAGGTCATAGGCTGAGGTCAGGCGCTCAAGCGGAGTTCGACCCCTACG 586
DB 524 ATCTCAAGGAAGTAATTTGGTGGGTAGATTTGTCAATGGTAGGAGTTATCCATACG 583
QY 587 ATGACGAGGACACGGAACCAAGTTCGGGGTATGTTGCCGGAACCGGACAGG---TTA 643
DB 584 ATGACCATGAGCATGGAACCTCATGAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTA 643
QY 644 ACTCCAGTACATAGGGTCTGCCCGGCGGAAGCTCTGCGGGTCAAGGTTCTCGGTTG 703
DB 644 ATGCAAGTTACAAGGAATGCTCCAGGAGCTAAGCTGCGGGAATTAAGGTTCTAGGTTG 703
QY 704 CGACAGGTTCCGGAAGGCTCTCCACCATCATCGGGGTGTTGACTGGGTGCTCGAAGAA 763
DB 704 CGATGGTCTTGGGAAGCATATCTACTATAATTAAGGGAGTTGAGTGGGCGGTTGATAA 763
QY 764 AGCAAGTACGGGATACGCTCATCAAGCTCTCCCTCGGCTCTCCAGAGTCTCCGAGG 823
DB 764 AAGTAAGTACGGAATTAAGTCAATTTCTTCTTGGTTCAAGCCAGAGTCTCCGAGG 823
QY 824 GAACGACTCCCTCAGTCAAGGCTCAAGAGCGCTGGAGCGCGGTATAGTACTGCTGG 883
DB 824 GAACGACTCCCTCAGTCAAGGCTCAAGAGCGCTGGAGCGCGGTATAGTACTGCTGG 883
QY 884 TCGCCGCGGCAACAGCGGCGGCAACCTACACCGCTCGGCTCACCCGCGCGCGGAGCA 943
DB 884 TCGCCGCGGCAACAGCGGCGGCGGCAACCTACACCGCTCGGCTCACCCGCGCGCGGAGCA 943
QY 944 AGTCTAACGCTGGTGTGAGTTGACAGCAAGCAACATCCAGCTTCTCCAGCAGG 1003
DB 944 AGTCTAACGCTGGTGTGAGTTGACAGCAAGCAACATCCAGCTTCTCCAGCAGG 1003
QY 1004 GACCGAGCGGAGCGGAAGCTTAAGCCGGAAGTGTGCTCCCGCGGCGGTTGACATCATAG 1063
DB 1004 GACCGAGCGGAGCGGAAGCTTAAGCCGGAAGTGTGCTCCCGCGGCGGTTGACATCATAG 1063
QY 1064 CCGCGCGGCGGAGCGGCAAGCTATGGGCAACCGCTCATATGAGACCGCGGAGATAGTCG 1123
DB 1064 CCGCGCGGCGGAGCGGCAAGCTATGGGCAACCGCTCATATGAGACCGCGGAGATAGTCG 1123
QY 1124 CTGGAAACAGCATGGCCACCCCTCACGTTTCGGGGTGTGGCGGCTCATCTCCAGGCC 1183
DB 1124 CTGGAAACAGCATGGCCACCCCTCACGTTTCGGGGTGTGGCGGCTCATCTCCAGGCC 1183
QY 1184 ACCGAGCTGGACCCCGGCAAGCTATGGGCAACCGCTCATATGAGACCGCGGAGATAGTCG 1243
DB 1184 ACCGAGCTGGACCCCGGCAAGCTATGGGCAACCGCTCATATGAGACCGCGGAGATAGTCG 1243
QY 1244 CCGCGAGGAGTACCGGACATCGCTACGGTGGGGTATAGGTTGAGGTTCTACAGGCCA 1303
DB 1244 CCGCGAGGAGTACCGGACATCGCTACGGTGGGGTATAGGTTGAGGTTCTACAGGCCA 1303

db	601	CTACTGGGACACGGCTCGAGGCACTCGACCTCTACCTCTACGACCCACCGGACGA	660
Qy	1473	GTTTGACTACTCTACACCGCCTACTAGGCTTCGAGAAGTTCGGCTACTACACCGAC	1532
Db	661	GTTTGACTACTCTACACCGCCTACTAGGCTTCGAGAAGTTCGGCTACTACACCGAC	720
Qy	1533	CGCGGAACCTGGACGGTCAAGGTCGTCACTCAAGGGCGGGCGAAGTACCAGTCCA	1592
Db	721	CGCGGAACCTGGACGGTCAAGTTCGTCACTCAAGGGCGGGCGAAGTACCAGTCCA	780
Qy	1593	CGTGTGAGGACGGAGGCGCTCAGCCAGTTCGGGGGGGGAACCGGAATCCAAACCCCAA	1652
Db	781	CGTGTGAGGACGGAGGCGCTCAGCCAGTTCGGGGGGGGAACCGGAATCCAAACCCCAA	840
Qy	1653	CCGGAACCCAAACCCGACACCGACCGAGACTTCCACGGTTCGGTTAACGACTACTG	1712
Db	841	CCGGAACCCAAACCCGACACCGACCGAGACTTCCACGGTTCGGTTAACGACTACTG	900
Qy	1713	GGACACACGGACACCTTCACCATGACGCTCAACAGGGTGCACGACGATACCGGTCA	1772
Db	901	GGACACACGGACACCTTCACCATGACGCTCAACAGGGTGCACGACGATACCGGTCA	960
Qy	1773	CCTGACCTTCGATACTTCTCTACACGACCTCGACCTTCTACCTCTACGACCCCAACCGCAA	1832
Db	961	CCTGACCTTCGATACTTCTCTACACGACCTCGACCTTCTACCTCTACGACCCCAACCGCAA	1020
Qy	1833	CCTCGTTGACAGTGCACAGCTCCAGCAACAGCTACGAGCAGCTCGAGTACGCGCAACCCCGC	1892
Db	1021	CCTCGTTGACAGTGCACAGCTCCAGCAACAGCTACGAGCAGCTCGAGTACGCGCAACCCCGC	1430
Qy	1893	CCCGGGAACCTGACAGTTCCTCGTCTACGCCCTACAGACCTACGCGCTGGGCGCACTACCA	1952
Db	1081	CCCGGGAACCTGACAGTTCCTCGTCTACGCCCTACAGACCTACGCGCTGGGCGCACTACCA	1140
Qy	1953	GCTCAAGCGCTGCTCTACTACGGG	1977
Db	1141	GCTCAAGCGCTGCTCTACTACGGG	1165

RESULT 5
 US-08-750-532-7
 Sequence 7, Application US/08750532
 Patent No. 5756339
 GENERAL INFORMATION:
 APPLICANT: MITTA, Masanori
 APPLICANT: YAMAMOTO, Katsuhiko
 APPLICANT: MORISHITA, Mio
 APPLICANT: ASADA, Kiyozo
 APPLICANT: TSUNASAWA, Susumu
 APPLICANT: KATO, Ikunoshin
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
 STREET: 419 Seventh Street N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: United States of America
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIMUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/750,532
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP95/01095
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894.818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34

Query Match 37.9%; Score 750.2; DB 4; Length 1962;
Best Local Similarity 64.7%; Pred. No. 1.9e-141;
Matches 1219; Conservative 0; Mismatches 618; Indels 48; Gaps 5;

QY 110 AGAAGAACTACGAGCTCTGACCCGCGACTGTTCAAGAAAGTCCAGAGGATGAAGTCTGA 159
DB 107 AGAAGAACTATGCTGCTCAAGCCAGAGCTGTTCAAGAAATTAAGTATGATCTCTA 166
QY 170 ACCAGGAATGGACAGCTCAATATGTTGGGAGCTAGGAGACAGGGACAGGGCGGTGA 229
DB 167 ACGAGGAATCAGCAGTAATGTTATTTGAAACCAATAGGGAAGAAATTCAGATTA 326
QY 230 AGGTACTGAGCTCATGGGCGCCAGGTCAAGTACTCTCAAGATAATCCCTGCTGTCG 285
DB 227 GAGTCTTGATTAATGGGTGCAGAAAGTTAGTATGTATACCATATTATACCCGCAATG 286
QY 290 CGGTTAAATTAAGGCGAGGACCTCTGCTGATCGGCGCATGATAGACAGGTTACT 343
DB 287 CTGCGGATCTTAAGGTATAGAGCTTACTAGTCTCAGTCTTAACAG----- 334
QY 350 TCGGTAAACAAAGGCTCTCGGCATAAAGTTTCATACAGAGGATTAACAAGTTCAGTT- 408
DB 335 GGGGTAAAGCTAAGCTTTACAGGTGTAGGTTTATCCAGGAGACTACAAGTTACAGTT 394
QY 409 -----GAGCAGCCACTTCGCTCTCCAGTAGGGCGGATACCGTTCTGGA 454
DB 395 CAGCAGAAATTAAGAGGACTGGATGAGTCTGCGAGCTCAAGTATGGCAACTTACGTTTGA 454
QY 455 ACTCCCTCGGTAGCAGGGAAGCGGTGTTGGTTCATCGATCGGTATAGGATAGACG 514
DB 455 ACT---TGGGATATGATGGTTCTGGAATCAACAATAGGAATTAAGACCTGGAATTGACG 511
QY 515 CGAACCCCGGATCTGAAGGCAAGGTCAATAGGTGTAGGCGCGGTCAACGCGAGT 574
DB 512 CTTCCTCATCCAGATCTCCAAGGAAGTAATTTGGTGGGTAGATTTTTCATATGTAGGA 571
QY 575 CGACCCCTACGATGACAGGAGACAGGAACCCAGCTTGCAGGTATCGTTGCCGAACCG 634

DB 572 GTTATCCATACGATGACCAATGACATGGAACTCATGTAGCTTCAATAGCAGCTGGTACTG 631
QY 635 GCAGCG---TTAACTCCCACTATAGCGCTGCCCGCGGCGGAAGCTCGTCCGGCTCA 691
DB 632 GAGCAGCAAGTAATGCAAGTCAAGGGAATGGCTCCAGGAGCTAAGCTGGCGGGAATTA 691
QY 692 AGGTTCTCGGTCCGACGAGTTCGGGAAGCGTCTCCACCATCATCGCGGTGTTGACTGG 751
DB 692 AGGTTCTAGTGGCGGATGGTTCGGAAGCATATCTACTATTAATTAAGGAGTTCAGTGG 751
QY 752 TCGTCAGAACAGCAAGCAAGTACGGGATAGGGTCAATCAACCTCTCCCTCGGCTCTCC 811
DB 752 CGGTTGATAACAAAGATAAGTACGGAAATTAAGSTCATTAATCTCTCTGTTGTTCAAG 811
QY 812 AGAGTCCGAGGGAACGAGCTCCCTCAGTCAGCGCTCAACAGCGCTGAGCGCGGTA 871
DB 812 AGAGTCAAGTCTACTGACGCTCTAAGTCAGCGCTTAAATGAGCGCTGGATGCTGGAT 871
QY 872 TASTAGTCTGGCTCGCGCGGCAACAGCGGCGGCAACACCTACACCTCGGCTCAGCG 931
DB 872 TAGTGTGTTGTTGCGCGCTCGAAGACAGTGGACCTTAACAGTATACAAATCGGTTCC 931
QY 932 CGCGCGGAGCAAGGTCAATACCGTCCGTCAGTTCACAGCAACGACCAATCGCCAGCT 991
DB 932 CAGCTGCAAGCAAGTATTAAGAGTTCGAGCGCTGACAGTATGATGTTATTAAGAGCT 991
QY 992 TCTCAGAGGAGGACCGACCGCGGCAACAGCGGCGGCAACACCTACACCTCGGCTCAG 1051
DB 992 TCTCAGAGGAGGCGCAATTCGACGCGGCGGCTTAAGCCCTGAGGTTGTTGCTCCAG 1051
QY 1052 TTGACATCATAGCCCGCGCTCCAGCGGGAACAGCATGCGCACCCCGATAAAGAGTACT 1111
DB 1052 ACTGGATTAATGCTCCGACAGCAAGTGAAGTATGATGCTCAACCAATTAATGACTAT 1111
QY 1112 ACACCAAGGCTCTTGGAACCAATGATGCGCACCGCGAGCTTTCGGGGGTTGGCGCGCTCA 1171
DB 1112 ACACAGAGCTCTCGGGAATCAATAGGCAACTCTCAGTAGTGGTATTCGACCGCTCT 1171
QY 1172 TCTCAGCGCCACCGGAGTGAACCGCGGAGAGGTGAAGACCGCGCTCATCGAGACG 1231
DB 1172 TGTCCAGCACACCGGAGCTGAGCTCCAGACAAAGTAAAAACAGCGCTCATAGAACTG 1231
QY 1232 CGGATAGTGGCGCCCGACAGTACGAGGAGATCGCTAGGTGGGTAGGTGAGG 1291
DB 1232 CTGATATGTAAGCGAGATCAATAGCGGATAGCTTACGCTGCGGAGGTGAGGTAA 1291
QY 1292 TCTAAGGCGCATCAAGTACGAGTACGCAAGCTCACCTTCACCGGCTCGCGCTCG 1351
DB 1292 CATACAAGGCTATAAGCTACGATAACTATGCAAGCTAGTGTTCACCTGGATATGTTG 1351
QY 1352 ACAGGGAAGCGCCACCACTTCGAGCTCAGCGCGCGCGCTTCGTCACCGCCACCC 1411
DB 1352 ACAGGCGCAACCACTCACCACTTCGTTATTAAGCGGAGCTTCGTTGCTACTGCCAT 1411
QY 1412 TCTACTGGSACACGGGCTCGAGTGAACATCGACTCTACTCTACGACCCCAAGCGGAAG 1471
DB 1412 TATACTGGGACAATGCCAATAGTGAAGCTTGTATCTCTACGATCCCAATGGAACCC 1471
QY 1472 AGGTTGACTCTCTACACCGCTACTACGGTTCGAGAGGTGGGTACTACAAACCGGA 1531
DB 1472 AGGTTGACTCTCTACACCGCTACTATGGATTCGAAAGGTTGGTTATACAAACCGGA 1531
QY 1532 CGCGCGGAACCTGGAGCGGTCAAGGTGTCAGCTACAGGCGCGCGGCACTACGAGGTG 1591
DB 1532 CTGATGGAACATGGCAATTAAGGTTGTAAGTACAGCGGAGTCAACATATCAAGTAG 1591
QY 1592 ACCTGCTCAGGAGGAGGCTCAGCGAGTCCGCGGCGGCAACCGCAACCAACCCCA 1651
DB 1592 ATGTTGATGATGGTTCCCTTTCAGAGCTTGGAG-----TTCACCT 1636
QY 1652 ACCGGAACCAACCGCGGAGTCCGCGGAGCTTCACCGGTTCCGTTAAGGACTACT 1711

Db 1637 CTCACACAGCAGACACACAGTAGACGCAAGACGTTTCCAGGATCCGATCACTACTACT 1656
Qy 1712 GGGACACAGCAGCAGCTTACCATGAACGTCACAGCGTCCGACCAAGATACCGGTG 1771
Db 1697 ATGACAGGAGCAGACCTTTACATGACGTTTAACTCTGGGCTACAAGATTAAGTGAG 1756
Qy 1772 ACCTGACCTTCGATGACTTCTTACACAGCAGCTTACCTTACGACCCCAAGCGCA 1831
Db 1757 ACCTAGTGTGTGACACAGCTACCATGATCTTGACCTTTACCTCTACGATCTTAACAGA 1816
Qy 1832 ACCTGCTTACAGCTTCCAGCTCGAGCAACAGCTACGAGCAGCTCGAGTACGCCAACCCG 1891
Db 1817 AGCTTGTAGATAGATCGGAGATCCCNACAGCTACGACACGTAAGATTAACCCCG 1876
Qy 1892 CCGCGGACCTGGAGCTTCTGCTTACGCTTACGCTTACGACCTACGCTTGGGGGACTAC 1951
Db 1877 CCGCGGACCTGGAGCTTCTGCTTACGCTTACGCTTACGCTTACGCTTGGGGGACTAC 1936
Qy 1952 AGCTCAAGCCGCTGCTACTACG 1976
Db 1937 AGCTGACGCTTAAGTTTATTATG 1961

RESULT 8

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Query Match 37.9%; Score 750.2; DB 4; Length 1962;
Best Local Similarity 64.7%; Pred. No. 1.9e-141;
Matches 1219; Conservative 0; Mismatches 618; Indels 48; Gaps 5;
Qy 110 AGAAGAACTACGAGCTGCTGACCCGCGGACTGTTCAAGAAAGTCCAGAGGATGACGGA 169
Db 107 AGAAGAACTATGCTCTGCTTAACCCGAGCTGTTTCAGAAATTCAGAAATTCAGCTTA 166
Qy 170 ACCAGGAAGTGGACACCGTCATATGTTGGGAGCTACGAGACAGGAGGAGGCTTA 229
Db 167 ACAGGAATACGACAGAGTAAATGTTGAAACCATAGGAAAGAAATTCAGCTTA 226
Qy 230 AGGTACTCAGGCTCATGGGCGCCAGGTCAAGTACTCCTTACAAGATTAATCCCTGCTG 289
Db 227 GAGTCTTGTAGTTAATGGGTGCAAAAGTTAGTATGTGATCATATATACCCGCAATAG 286
Qy 290 CCGTTAAATTAAGGCGGAGGACCTTCTGCTGATCGGGGCTATGATAGACAGGCTTACT 349
Db 287 CTGGCGATCTTAAGGTTAGAGACTTACTAGTCACTCTCAGGTTTAAACAG----- 334
Qy 350 TCGGTACACAGAGGCTCTGGGCGATAAAGTTCATACAGGAGGATTAAGAGTTTCAAGTT 408
Db 335 GGGGTAAAGCTAAGCTTTCAGGTGTTAGGTTTATCCAGGAAGACTACAAGTTTACAGTT 394

Qy 409 -----GACGAGCCAGCTTCCGCTCTCCAGATAGGGCCGATACCGCTCTGCA 454
Db 395 CAGCAGAAATTAGAAGGACTTGGATGAGTCTGAGCTCAAGTTATGGCAACTTACGTTTGA 454
Qy 435 ACTCCCTCGGCTACGAGCAAGCGGTGTTGCTGCTGATACGGGTATAGAG 514
Db 455 ACT---TGGGATATGATGCTCTGGAATCACAAATAGGAATTAATTGACACTGGAATTGACG 511
Qy 515 CGAACCCACCCGATCTGAGGCAAGGTCATAGCTGGTACGACCGCTCAACGCGAGT 574
Db 512 CTCTTCATCCAGATCTCCAAAGGAAAGTAATTGGGTGGGTAGATTTTGTCAATGCTAGGA 571
Qy 575 CGACCCCTACGATGACGAGGACGACGACGCTTGGGGTATCGTTTGGCGGAACCG 634
Db 572 GTTATCCATACGATGACATGGACATGGAATCTATGCTTCAATAGCAGCTGTGACTG 631
Qy 635 GAGCG---TTAACTCTTAACTAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGT 691
Db 632 GAGCAGCAAGTAAATGGCAAGTACAAAGGAAATGGCTCCAGGAGCTAAGCTGGCGGAATTA 691
Qy 692 AGGTTCTCGGTGCGGAGCTTGGGAGGCTTCCACCATATCGGCGGTGTTGACTGG 751
Db 692 AGGTTCTAGGTGCGGATGTTCTGGAAGCATATCTACTATTAATTAAGGAGTTGACTGG 751
Qy 752 TCGTCCAGAAACAGCAAGTACGAGGATAGGATCATCAACCTCTCTCCCTGGCTCTCTCC 811
Db 752 CCGTTGATTAACAAAGATTAAGTACGAGGCTCAACAGGCTTCTCTCTCTCTCTCTCTCTCT 811
Qy 812 AGAGCTCGGAGGAGGAGCTTCCCTCAGTCAAGGCGTCAACAGGCTGGGAGCGGTA 871
Db 812 AGAGCTCGGAGGAGGAGCTTCCCTCAGTCAAGGCGTCAACAGGCTGGGAGCGGTA 871
Qy 872 TAGTATGCTGCTGCGGCGGCGCAAGAGGCGGCAACACTACACGCTCGGCTCACCG 931
Db 872 TAGTATGCTGCTGCGGCGGCGCAAGAGTGGAGCTTAAAGTATACAACTGCTCTCTCTCT 931
Qy 932 CCGCGCGGAGCAGCTCATAACTGCTGGGTGAGTGTGACAGCAACAGCTCGGCTGAGT 991
Db 932 CAGCTGCAAGCAAGTTATACAGTTGGAGCGTTGACAAAGTATGATGTTTATAACAGCT 991
Qy 992 TCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1051
Db 992 TCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1051
Qy 1052 TTGACATCATAGCCCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1111
Db 1052 ACTGATATAATGCTGCGAGAGCAAGTGGAACTAGCATGGGTCAACCAATTAATGACTATT 1111
Qy 1112 ACACCAAGCCCTCTGGAACGAGTGGCCACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1171
Db 1112 ACACAGAGCTCTCTGGGAGCATCAATGGCACTCTCAGTACGTTAGTGGTTCAGGCGCTCT 1171
Qy 1172 TCTCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1231
Db 1172 TCTCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1231
Qy 1232 CCGAGATAGTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1291
Db 1232 CTGATATGTTAAAGCCAGATGAAATAGCCGATATAGCTACGCTGAGGAGGAGGAGGAGG 1291
Qy 1292 TCTACAGGCGCATCAATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1351
Db 1292 CATACAGGCTTAAACTAGCTTAACTAGCTTAACTAGCTTAACTAGCTTAACTAGCTTAA 1351
Qy 1352 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1411
Db 1352 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1411
Qy 1412 TCTACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1471
Db 1412 TATACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1471
Qy 1472 AGGTTGACTACT 1531

Db 1472 AGGTGACTACTTACACCGCTACTATGATTCGAAAGGTGGTTATACACCCAA 1531
Qy 1532 CCACCGGAACCTGGAGGTCAGGTCAGCTACAGGCGGCGGAGCACTACAGGTGC 1591
Db 1532 CTGATGACATGAGCAATTAAGTTGTAAGCTACAGCGGAGTGCAAACTATCAAGTAG 1591
Qy 1592 ACCTGCTCAGCAGCGGAGCTCAGCCAGTCCGCGCGGCAACCGAATCCAAACCCCA 1651
Db 1592 ATGTGGTAGTGTATGTTCCCTTTCACAGCGCTGAG-----TTACCAT 1636
Qy 1652 ACCGGAACCCACCGGACCGGACCGGACCGGACCGGTCACCGGTTCCGTTAACGACTACT 1711
Db 1637 CTCACAAACCAACAGCTAGAGCAAGAGCGTTCCAAAGGATCCGATCACTACTACT 1696
Qy 1712 GGGACACAGCAGCAGCTTACCATGAGCTGCAACAGCGGTGCAACAGATACACCGGTG 1771
Db 1697 ATGACAGGAGGACACCTTTACATGACCGTTACTCTGGGCTACAAAGATTACTGGAG 1756
Qy 1772 ACCTGAGCTTCGATCTCTACACAGCAGCTGACCTCTACCTCTACGACCGCAAGCGCA 1831
Db 1757 ACCTAGTGTGTTGACACAGCTACCATGATCTTGACCTTTACCTCTACGATCCTAACCGA 1816
Qy 1832 ACCTGCTGACAGCTCCAGCTCGACCAACAGCTACGAGCAGCTGCGGACCGCAACCGCG 1891
Db 1817 AGCTTGTAGATAGTGGAGAGTCCCAACAGCTAGCAACAGCTAGAAATCTTAACCCCG 1876
Qy 1892 CCCCGGAGCTGGAGGTTCTGCTGCTAGGCTACAGCAGCTACGCTGGGCGGACTACC 1951
Db 1877 CCCGAGAGCTGGTACTTCTCTAGTATATGCTACTACTACTACGCTTGGGCTTACTACG 1936
Qy 1952 AGCTCAGGCGCTGCTACTACGG 1976
Db 1937 AGCTCAGCGCTTAAGTTTATTATGG 1961

RESULT 9

US-08-894-818B-4
Sequence 4, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, HIKARU
APPLICANT: MORISHITA, MIO
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: HIRATA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894.818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/01253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,518
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
OTHER INFORMATION: /note- N at position 1283 is G or T.
US-08-894-818B-4

Query Match 35.0%; Score 692.8; DB 4; Length 1566;
Best Local Similarity 67.1%; Pred No. 5.3e-130;

Matches 1020; Conservative 0; Mismatches 483; Indels 18; Gaps 2;

Qy 459 CTTGCGGTACAGCGGAGCGGTGTGGTGTCCCATCGTCTCGATACGGGTATACACCGGAA 518
Db 60 CTTGGGATATGATGTTCTGGATCAATAGGAATAATTGACACATCGAATTGACGCTTC 119
Qy 519 CCACCCCGATCTGAAGGCTAAGGTCATAGGCTGGTACGACCGCTCAACGCGAGCTCGAC 578
Db 120 TCATCCAGATCTCCAAAGGAATGATTTGGTGGCTAGATTTTGTCAATGTAGGAGTTA 179
Qy 579 CCGCTTCGATACAGCGGACAGCGGACGACCGGAGTTCGCGGTATCGTTCGCGGAGACCGGAC 638
Db 180 TCCATACGATGACCTGGTACATGGAATGGAATCTCATGTAGCTTCAATAGCAGCTGGTAC 239
Qy 639 CG---TTAACTCCCATACATAGGCTTCGCGCGCGGAGAGCTCGTTCGCGGCTCAAGGT 695
Db 240 AGCAAGTATGCAAGTACAGGGAATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGT 299
Qy 696 TCTCGGTCGCGGAGTTCGCGGAGCGTCTCCACCATCATCGCGGGTGTGTACTGGGTCTGT 755
Db 300 TCTAGTCCGATGTTCTGCAAGCATATCTACTATAATTAAGGAGGTGTAGTGGGCGGT 359
Qy 756 CCAGAACAGGACAAAGTACGGGATTAAGGTCATCAACCTCTCCCTCGGCTCCCTCCACAG 815
Db 360 TGATACAAAGATAAGTAGGCAATTAAGGTCTAATCTTCTCTGTTGTTCAAGCCAGAG 419
Qy 816 CTCGAGCGGACCGGACCTCCCTCAGTACGCGGCTCAACACGCGTGGGAGCGGCTATAGT 875
Db 420 CTCAGATGGTACTGACGCTCTAAGTCAAGCTGTTAATGACGCGTGGGATGCTGGATTAGT 479
Qy 876 AGTCTGCGTCCGCGCGGACAGCGGCGGACACCTACACCGCTCGGCTCACCGCGCGC 935
Db 480 TGTGTGGTTGCCGCTGGAAFCJTGACCTAACAAAGTATACATCGGTTCTCCAGCAGC 539
Qy 936 CGGAGCAAGTCTAACCTCTGTCAGTTGACAGCAAGCAACATCGCGGCTCTC 995
Db 540 TGCAGCAAGTATTAAGCTGTGGGCGGTGACAAAGTATGATTTATACAAAGCTTCTC 599
Qy 996 CAGCAGGCGGACCGCGGAGCGGAGGCTCAAGCGGAGTCTGCGCCCGCGGCTTGA 1055
Db 600 AAGCAGAGGGCCAACTGACAGCGGAGGCTTAAGCGTGTGCTCCAGGAACTG 659
Qy 1056 CATCATGCGCGCGGCGGAGCGGAGGAGTGGGACCGGAGTGGGAGTGGGAGTGGGAGTGG 1115
Db 560 GATAATGCTGCGAGGAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 719
Qy 1116 CAAGGCGCTCTGGAACAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 1175
Db 720 AGCAGCTCTCTGGGAGTCAATGGCAACTCTCAGCTAGTGGTATTTGCGGCTCTTGGCT 779
Qy 1176 CAGGCGCGGAGCTGAGCGCGGAGAGGAGTGAAGCGGCTCATCGGAGCGGCGCA 1235
Db 780 CCAGCAGACCGGAGCTGAGCTGAGCAAGTAAAGCAAGTAAAGCAAGCTATAGAACTGCTGA 839

QY 1236 CATAGTCGCCCCCAAGAGATAGCGACATCGCTACGGTGGGGTAGGGTGAACGCTA 1295
DB 840 TATCGTAAACAGAGATGAATAGCGATATAGCGTACGGTGAAGTGAAGTGAATGCA 899
QY 1296 CAAGGCCATCAAGTACAGAGATAGCGACATCGCTACCGGCTCCGTCGGGACAA 1355
DB 900 CAAGGCTATAAATACAGATGAATATGCAAGCTAGTGTCTCACTGAGATATGTTGCCAACAA 959
QY 1356 GGAAGAGCCACACACACCTTCGAGCTCAGCGCGCCACCTTCGTCAGCCGCCACCTCTA 1415
DB 950 AGGACGCCAAACTCACCAGTTCGTTATTAGCGAGCTTCGTTTCTGAATGCCACATTATA 1019
QY 1416 CTGGAGACCGGGTCCAGGAGATCGAGCTCTACCTCTAGACGCCCAAGCGGAGAGGT 1475
DB 1020 CTGGAGACATGCGCAATAGCGAGCTTGATCTTTACCTCTAGGATCCCAATGGAAGCAGGT 1079
QY 1476 TGACTACTCTACACCGCTTACTAGCGCTTCGAGAGGTGGGTACTACAAACCGACCGC 1535
DB 1080 TGACTACTCTTACACCGCTTACTAGGATTCGAAAGGTGGTTATTACAAACCACTGA 1139
QY 1536 CGGAACCTCGAGCTCAAGTCTCAGCTACAGCGCGCGGCGCAACTACAGTCTGAGGT 1595
DB 1140 TGAACATGAGCAATTAAGTGTAGCTACAGCGAGGTGCAAGTATCAAGTATAGTGT 1199
QY 1596 GGTGAGCGAGGAGCTTCAAGCTTCCGCGCGCGGCAACCCGAAATCCAAACCCCAACCC 1655
DB 1200 GGTAGTGTGCTTCCCTTTCACAGCTGGAAG-----TTCACCATCTCC 1244
QY 1656 GAACCCACCGCCAGCACACCGACCTTCACCGCTTCGTTAAGCACTACTGGA 1715
DB 1245 ACACACAGACCAAGAGTACAGCAAGAGCTTCAGAGTTCGATCTACTATATGA 1304
QY 1716 CACGAGCGACACTTCAAGTACAGCTCAAGCGGTGCCACCAAGATACCGGTGACT 1775
DB 1305 CAGGAGCGACACTTTACATGACCTTAACTCTGGGCTACAAAGATTAAGTGAAGCT 1364
QY 1776 GACCTTCGATCTCTACAGCAAGCTGACCTCTACCTCTAGCGCCCAAGCAAGCT 1835
DB 1365 AGTGTGACACAGCTACCATGATCTTGACCTTTACCTCTAGGATCTTACCAAGCT 1424
QY 1836 GGTGACAGGTTCAGTTCAGCAAGCTACAGGAGCTGAGTACGCAAGCGGCGGCGCC 1895
DB 1425 TGATAGATAGTTCGAGAGTCCCAAGAGCTACGACAGCTAGATTAAGTTCAGCGCGCC 1484
QY 1895 GGAACCTTCGAGCTTCTCTCTACCGCTTACAGCAAGCTTACGCTTGGGCGGAGTACAGCT 1955
DB 1485 AGGAACCTGCTTCTCTCTAGTATGCTTACTACTACTACTAGCTTGGGCTTACTAGAGCT 1544
QY 1956 CAGGCGCTGCTTACTAGG 1976
DB 1545 CAGGCTTAAAGTTTATTATGG 1565

RESULT 10

US-09-445-472-2

; Sequence 2, Application US/09445472

; Patent No. 6358726

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/09/445,472

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 1236

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

Query Match

28.3%; Score 560.4; DB 4; Length 1236;

Best Local Similarity 68.0%; Pred. No. 1.4e-103;

Matches 796; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

QY 459 CTTCCGCTTACACGCAAGCGTGTGGTGTCCATCGCTGATACGGGTATACACCGCAA 518
DB 60 CTTGGATATATGTTCTTGGATACATAGGAATATTGACACTGGAATTGACGCTTC 119
QY 519 CCACCCCBATGTGAAGGCAAGGTATAGGCTGGGTACGACGCCGTCAACGGCAGGTGCGAC 578
DB 120 TCATCCAGTCTCCAAGCAAAAGTAATTGGGTGGGTAGATTGTTCAATGGTAGGAGTTA 179
QY 579 CCCCCTACGATGACGAGGACGCAACCCAGCTTGGGGTATCGTTGCCGGAACCCGCGAG 638
DB 180 TCCATACGATGACCATGGAATGGAATCATGTAGCTTCAATAGCAGCTGTACTGAGCG 239
QY 639 CG---TTAACTCCGAGTACATAGGCTGCGCCCGCGGCGGAGCTGTCGGGCTCAAGGT 695
DB 240 ACAAGTAAATGGCAAGTACAGGGAATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGGT 799
QY 696 TCTCGGCTCCGACCGTTCGGGAAGCGTCTCCACCATCATCGCGGCTGTGACTGGGTGCT 755
DB 300 TCTAGGTCCGATGCTCTGGAAGCATATCTACTATAATTAAGGAGTTCAGTGGCGCT 359
QY 756 CAGAACAGAGCAACTACGGTATAAGGCTCATCAACCTCTCCCTCGGCTCTCCCGAGAG 815
DB 360 TGATAACAAGATAGTACGGAATTAAGGTCTAATCTTCTCTCTTCAAGCCAGAG 419
QY 816 CTCGACGCAAGCACTCCCTCAGTCAGCGCGCTCAACACCCCTGGACCCCGGTATAGT 875
DB 420 CTCAGATGCTACTGAGCTCTAAGTCAAGCTGTAAATGACGGTGGGATGCTGATAGT 479
QY 876 AGTCTGCTGCGCGCGCAACAGCGGCGGAGACCTACACCTTCGCTCAGCTCAGCCGCGC 935
DB 480 TGTGTGTGCGCGCTGGAACAGTGGACCTAACAAGTATACAATCGGTTCTCCAGCAGC 539
QY 936 CGGAGCAAGTCTATACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 995
DB 540 TGAACCAAGTATTATGAGTGGAGCGGTGACAAAGTATGATGTTATAACAAGCTTCTC 599
QY 996 CAGCAGGCGCGCGCGCAAGGCTCAAGCGGAGTCTGCGCGCGCGCGCGCGCGCTTGA 1055
DB 600 AAGCAGAGCGCGCGCAAGGCTCAAGCGGAGTCTGAGCTGAGTGTGCTCCAGGAACTG 659
QY 1056 CATCATAGCGCGCGCGCAAGGCTCAAGCGGAGTCTGAGCTGAGTGTGCTCCAGCAGCT 1115
DB 660 GATAATGCTGCGAGAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 719
QY 1116 CAAGGCTCTGGAACACATAGCGCTACGGTACGGTGGGCTTGGCGGCTCATCCCT 1175
DB 720 AGCAGCTCTGGGAGATCAATGCGAACTCTCAGCTAGTGTGATTTGAGCGCTCTTCTCT 779
QY 1176 CCAGGCGCGCGCGCGCAAGGCTCAAGCGGAGTCTGAGCGCGCGCGCGCGCGCGCGCA 1235
DB 780 CCAAGCAGCGCGCGCGCAAGGCTCAAGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 839
QY 1236 CATAGTCCGCGCGCGCAAGGAGTACGGCTACGGTACGGTGGGCTTGGCGGCTCATCCCT 1295
DB 840 TATCGTAAAGCCAGATGAATAGCGATATAGCTACGGTGCAGGTAGGTTAATGATA 899
QY 1296 CAGGCGCATCAAGTACGAGGAGTACGCGCAAGCTCAGCTTTCAGCGGCTCGGTCGCGCA 1355
DB 900 CAAGGCTATAAATACAGTAACTATGCAAAAGTAGTGTTCAGCTGAGTATGTTGCCAA 959
QY 1356 GGAAGCGCGCGCGCGCAAGCTTTCAGCTCAGCGCGCGCGCGCGCGCGCTTCTGAGCGCGCGCGCTCTA 1415
DB 960 AGGACGCCAACTCAGGAGTTCGTTATTAGCGGAGCTTCGTTTCTGTAAGTGCACATTATA 1019

02-1775

QY 1416 CTGGGACACGGGCTCGAGGACATGACCTCTACTCTAGACCCCAAGGGAACGAGGT 1475
Db 1020 CTGGGACAAATGCCAATAGGACCTTGATCTTACCTCTAGCTCCCAATGGAAACGAGGT 1079
QY 1476 TGACTACTCTTACACCGCTTACTACGGTTCGAGAAGCTCGGCTACTACAAACCCGACCC 1535
Db 1080 TGACTACTCTTACACCGCTTACTACGGTTCGAGAAGCTCGGCTACTACAAACCCGACCC 1535
QY 1536 CGGAACCTGACCGGTCAAGCTGCTAGCTACAAAGGCGCGGCAACTACAGTTCGAGT 1595
Db 1140 TGAACATAGGAAATTAAGGTGTGATCTACAGCGGAAGTGCRAACTACTCAAGTAGTGT 1199
QY 1596 COTCAGCAGCGGAGCCTCAGCCAGTCCGG 1625
Db 1200 GGTAGTGATGTTCCCTTTTCACAGCCTGG 1229

RESULT 11
US-08-750-532-11
; Sequence 11, Application US/08750532
; Patent No. 5756339
; GENERAL INFORMATION:
; APPLICANT: MITTA, Masanori
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MORISHITA, MIO
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NETMARK, P.L.L.C.
; STREET: 419 Seventh Street N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,532
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/01095
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1994/130236
; FILING DATE: 13-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1994/173912
; FILING DATE: 26-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: MITTA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..564

US-08-750-532-11

Query Match 26.9%, Score 532.8, DB 1; Length 564;
Best Local Similarity 97.0%; Pred. No. 3.9e-96;
Matches 543; Conservative: 0; Mismatches 17; Indels 0; Gaps 0;
QY 598 CACGGAACCCAGCTTGGGGGTA*CGTTGGCGGAACCGGACGGTTAACTCCCAAGTACATA 557
Db 4 CAGGGAACCTCAGTGGGGGAACAGTTGCGGAACAGGACGGTTAACTCCCAAGTACATA 63
QY 658 GCGCTGCGCCCGCGCGGTA*AGTTCGTCGGGTCAAGGTTCCTGGTCCGACGGTTCGGGA 717
Db 64 GCGCTGCGCCCGCGCGGGAAGCTCGTCGGGTGCAAGGTTCCTGGTCCGACGGTTCGGGA 123
QY 718 AGCTCTCCACCATCATCGCGGGTGTGTGACTGGGTCTCCAGAACAGGACAAAGTACGGG 777
Db 124 AGCTCTCCACCATCATCGCGGGTGTGTGACTGGGTCTCCAGAACAGGACAAAGTACGGG 183
QY 778 ATAGGCTCATCAACCTCTGCTCGGCTCTCCAGAGCTCCGACGGACCGACTCCCTC 837
Db 184 ATAGGCTCATCAACCTCTGCTCGGCTCTCCAGAGCTCCGACGGACCGACTCCCTC 243
QY 838 AGTACGGCGTCAACACAGCT*GGACCGCGGTATAGTCTGCGTCCGCGCGGCAAC 897
Db 244 AGTACGGCGTCAACACAGCT*GGACCGCGGTATAGTCTGCGTCCGCGCGGCAAC 303
QY 898 AGCGGCGCAACACTACACGCTCGGCTACCGCGCGGTATAGTCTGCGTCCGCGCGGCAAC 957
Db 304 AGCGGCGCAACACTACACGCTCGGCTACCGCGCGGTATAGTCTGCGTCCGCGCGGCAAC 363
QY 958 GTTCAGTTGACAGCAACGA*ACATCGCGAGCTTCTCCAGCGGGACCGCGCGGAC 1017
Db 364 GTTCAGTTGACAGCAACGA*ACATCGCGAGCTTCTCCAGCGGGACCGCGCGGAC 423
QY 1018 GGAAGCTCAACCGGAGTGTGTCGCCCGCGGTGACATCATAGCCCGCGCGGCGAC 1077
Db 424 GGAAGCTCAACCGGAGTGTGTCGCCCGCGGTGACATCATAGCCCGCGCGGCGAC 483
QY 1078 GGAACGAGTGGGCGACCGCGGTAAAGAGTACTACTACCAAGGCTCTGGACAGCATG 1137
Db 484 GGAACGAGTGGGCGACCGCGGTAAAGAGTACTACTACCAAGGCTCTGGACAGCATG 543
QY 1138 GCCACCGCGCACTGTCGG 1157
Db 544 GCCACCGCGCACTGTCGG 543

RESULT 12
US-08-894-818B-14
; Sequence 14, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hitaru
; APPLICANT: MORISHITA, MIO
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Netmark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/894,818B
;; FILING DATE: 20-MAY-1998
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP96/03253
;; FILING DATE: 07-NOV-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 321285/1995
;; FILING DATE: 12-DEC-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Browdy, Roger L.
;; REGISTRATION NUMBER: 25,618
;; REFERENCE/DOCKET NUMBER: TAKAKURA-1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 628-5197
;; TELEFAX: (202) 737-3528
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 564 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: other nucleic acid
;; US-08-894-818B-14

Query Match 26.9%; Score 532.8; DB 4; Length 564;
Best Local Similarity 97.0%; Pred. No. 3.9e-98;
Matches 543; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 598 CACGGACCCAGCTTGGGGTATGTTGGCGGAACCGGACGGTAACTCCCATACATA 567
DB 4 CACGGACCTACCTGGCGGGAACAGTTGCCGGAACAGGAGCGTTAACTCCCATACATA 63
OY 658 GCGCTCGCCCGCGGCGGAAGCTGTCGGCTCAAGGTTCTCGTCCGACGGTTCCGGA 717
DB 64 GCGCTCGCCCGCGGCGGAAGCTGTCGGCTCAAGGTTCTCGTCCGACGGTTCCGGA 123
OY 718 AGCGTCTCCACCATATCGCGGTGTTGACTGGTCTCCAGAACAGCAAGTACGGG 777
DB 124 AGCGTCTCCACCATATCGCGGTGTTGACTGGTCTCCAGAACAGCAAGTACGGG 183
OY 778 ATAAGGCTCATCACTCTCCCTCGGTCTCCAGAGTCTCCAGAACAGTCCCTC 837
DB 184 ATAAGGCTCATCACTCTCCCTCGGTCTCCAGAGTCTCCAGAACAGTCCCTC 243
OY 838 AGTCAGCGCTCAACAGCGCTGGAGCGGTATAGTCTGCTCGCGCGGCGAAC 897
DB 244 AGTCAGCGCTCAACAGCGCTGGAGCGGTATAGTCTGCTCGCGCGGCGAAC 303
OY 998 AGCGGCGGCAACCTACACCTCGGTCTACCCCGCGCGGAGCAAGTCTAATC 957
DB 304 AGCGGCGGCAACCTACACCTCGGTCTACCCCGCGCGGAGCAAGTCTAATC 363
OY 958 GGTGCACTTGACAGCAACAGCAACATCGCAGCTTCTCAGAGGGGACCGCGGAC 1017
DB 364 GGTGCACTTGACAGCAACAGCAACATCGCAGCTTCTCAGAGGGGACCGCGGAC 423
OY 1018 GGAAGGCTCAGCGGAGTGTGTCGCGCGCGGCTTACATCATAGCCCGCGCGGACG 1077
DB 424 GGAAGGCTCAGCGGAGTGTGTCGCGCGCGGCTTACATCATAGCCCGCGCGGACG 483
OY 1078 GGAACAGGATGGGACCGCCGATAAAGCACTACTACACCAAGGCTCTGGAACAGCATG 1137
DB 484 GGAACAGGATGGGACCGCCGATAAAGCACTACTACACCAAGGCTCTGGAACAGCATG 543
OY 1138 GCCACCCCGGACGTTTCGGG 1157
DB 544 GCCACCCCGGATGTTACCGG 563

RESULT 13
US-09-000-016-1

;; Sequence 1, Application US/09000016
;; Patent No. 6143541
;; GENERAL INFORMATION:
;; APPLICANT: Akira ARISAWA et al.
;; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
;; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
;; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
;; STREET: 2033 K Street, N.W., #800
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20006
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/000,016
;; FILING DATE: January 30, 1998
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warren M. Cheek, Jr.
;; REGISTRATION NUMBER: 33,367
;; REFERENCE/DOCKET NUMBER:
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-721-8203
;; TELEFAX: 202-721-8250
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2809 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: genomic DNA
;; ORIGINAL SOURCE:
;; ORGANISM: Streptomyces viridosporus
;; STRAIN: A-914
;; ORIGINAL SOURCE:
;; ORGANISM: Streptomyces antibioticus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 338...2535
;; IDENTIFICATION METHOD: 2
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2540...2809
;; IDENTIFICATION METHOD: 2
;; US-09-000-016-1

Query Match 7.7%; Score 191.6; DB 3; Length 2809;
Best Local Similarity 50.5%; Pred. No. 7.1e-30;
Matches 563; Conservative 0; Mismatches 439; Indels 48; Gaps 8;
OY 332 TGATAGACACGGGTACTTGGTTAACACACAGGCTCGGGCATAAAGTTTCATACAGAGG 391
DB 870 TGTGGAGCGGCTCACCACCGGACCGGCTCGGGCATCGCCCGCTGCTGGTGG 929
OY 392 ATTACAGGTTCAAGTTGAAGCGGACTTCCGCTCCAGATAGGGCGCGATACCGTCT 451
DB 930 ACGGGGTCCGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 989
OY 452 GGAACCTCCCTCGGCTACGAGGAAAGCGGTGTGGTGTGGCATCGATACGGGTATAG 511
DB 990 GG---TCCGCGGCTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1046

512 ACGGGAACACCCGATCTGAAGGCAAGAGTATAGCTGGTACAGCCGCTCAACGCA 571
1047 ACAGGAGCATCGGACCTGAAGGCGGGTGA---CCGGTCCAAGAACTTACCGCG 1103
572 GGTGACCCCTACGATGACGAGGACAGCAACCCACGTTGGGTATCGTTCGGGAA 631
1104 GCGCGCGCGCGGCGGAGGTCAGGTCAGGACCCACGTCGCTCGATTCGGCGGCA 1163
632 CCGGCAGC---GTTAACTCCCACTACATAGCGCTCGCGCGCGGCGGAGTCTCGCG 688
1164 CCGGCGCGGAGTCCAAAGGCAAGTACAAGGCGCTCGCGCGCGGCGGATCTCAACG 1223
689 TCAAGGTCCTGCTGGGCGGAGGTCGGAAGCGCTCCACATCATCGCGGGTGTGACT 748
1224 GCAAGGTCCTGAGGAGTCTGGGTCGGGAGGACTCGGATCTCGCGGCGATGAGT 1283
749 GGGTCTCCAGAACAAAGGACAAAGTACGGGATAGGGTCTACACCTTCCTCGGCTCT 808
1284 GGG-----CGGCGCGGAGGCGCGGAGCTCGTCAACATGAGCCTGCGCGCA 1331
809 CCGAGGTCGCGAGGCAAGCACTCCCTCACTCAGGCGCTCAACAAAGCGCTGGACGCG 868
1332 TGGACACCGGAGACCGACCGCTGAGGCGGCGGTGACAGCTGTCCGCGGAGAG 1391
869 GTATAGTATGCTGCTGCGCGCGGCGGAGACAGCGGCGGCGGAGACCTTACAGCTCGCTCAC 928
1392 GCGTCTGCTGCGGAGTCTGGGTCGGGAGGACTCGGATCTCGCGGCGGATGAGT 1449
929 CCGCGCGCGGAGGCAAGGTCATACGCTCGGTGAGGCGGCGGTGACAGCTGTCCGCGGAGAG 988
1446 CCGGAGCGGAGGCGCGGCTCACCGCTCGGCGGCGGCTGACGACAGGACAGCTCGCG 1505
989 GCTTCTCAGGAGGAGGAGCAGC---CGGAGAGGAGGCTCAAGCGGAGTCTCGCTCGCC 1045
1506 ACTTCTCTCCAGCGCGCGCGCTCGGCGAGCGGCGGATCAAGCGGAGCTCACCGCTC 1505
1046 CCGGCGTGTACATCATGACCGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1625
1566 CCGGCGTGTACATCATGACCGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1625
1101 ----AAAGGAGTATACAGAGGCTCTGGAACAGCATGCGGACCGGAGGAGGAGGAG 1156
1626 AGGAGCGCGCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1685
1157 GCGTGGCGGCTCATCTCCAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
1686 GCGCGCGGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1745
1217 CCTCATGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1276
1746 GCGTACCGGCTCCACGAGGCGG---AAGTACACCGGCTTCGAGCAGGCT 1796
1277 CCGGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1336
1797 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1856
1337 CCGGCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1386
1857 TGAGCTTCGGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1906

RESULT 14

US-09-514-340-1

Sequence 1, Application US/09514340

Patent No. 6361987

GENERAL INFORMATION:

APPLICANT: Akita ARISAWA et al.

TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRINDINE DE

ITS EXPRESSION PRODUCT

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/514,340

FILING DATE: 29-Feb-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/000,016

FILING DATE: January 30, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Lee Cheng

REGISTRATION NUMBER: 40,949

REFERENCE/DOCKET NUMBER: <Unknown>

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2809 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: Streptomyces antibioticus

STRAIN: <Unknown>

FEATURE:

NAME/KEY: CDS

LOCATION: 338...2539

IDENTIFICATION METHOD: E

FEATURE:

NAME/KEY: CDS

LOCATION: 2540...2809

IDENTIFICATION METHOD: P

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-514-340-1

Query Match 9.7%; Score 191.6; DB 4; Length 2809;
Best Local Similarity 54.5%; Pred. No. 7.1e-30;
Matches 583; Conservative 0; Mismatches 439; Indels 48; Gaps 8;

Qy 332 TGATAGACACGGGTACTTCGTAACACAAAGGTCCTCGGGCATAAAGTTTCATACAGGAGG 391
Db 870 TGTGGAGCGGCTCACCACAGGCGACGGACGCGCTCGGCATCGCCACGCTGTGGCTGG 929
Qy 392 ATTACAGGTTACGTTGAGTACGCGACCTTCGCTCCAGATAGAGGGCGGATACGCTGT 451
Db 930 AGGGGTTCGAGGCGCGGCTTCACACGCTCGTCCGACATCGCGCCCGCCCAAGGCT 989
Qy 452 GGAACCTCCCTCGGTAGCAGGAGGCGGTGTGTTGCCATCGTCGATACGCGGTATAG 511
Db 990 GG---TCCGCGGCTACGTCGCGTAGGATCGCGCTCTGGACACCGGTGTG 1046
Qy 512 ACAGGACACCGCGATCTAGGCGAAGGTCATAGGCTGCTAGGCGCGCTCAACGCA 571
Db 1047 ACAGGACACCGCGATCTAGGCGGCGGCTGAGGGCGGCTGAG---CGCGCTCAAGAACTTACCGCG 1103
Qy 572 GTTCGACCGCTTACGATGTCAGGAGACAGCAACCCAGCTTCGCGGTATCGTTCGCGAA 631
Db 1104 CCGCGCGGCGGCGGAGCAAAATGCGGCGAGGACCGACCGCTCGATCGCGCGGCA 1163
Qy 632 CCGGCGAGC---GTTAATCTCAGTACATAGGCGCTCGCGCGGCGGAGCTGTGCGG 688

Db 1164 CGGGCGCCAGTCCAGGCGAAGTACAGGGGGTCCACCGGGCGCGGATCCTCAAG 1233
Qy 589 TCAAGTTCTCGGTGCGAGGTTTCGGAAGGCTCTCCACCATCATCGGGGTTGACT 748
Db 1224 SCAGGTCTCGAGACTTCGGTTTCGGGAGGACTCGGCTCTCCCGGCGATGGAGT 1283
Qy 749 GGGTCTCCAGACAGGAGAGTACGGGTATAGGGTATCAACCTCTCCCTGGCTCT 808
Db 1284 GGG-----CGGCGCGCAGGGGCGCGAGCTCTCAACATGAGCTGGCGGCA 1331
Qy 809 CCCAGAGTCCGAGCGGAACGAGTCTCAGTCAGGCGGTCAACAAGCGCTGGAGCGG 868
Db 1332 TGGACACACCGAGAGCGCCGCTGGAGGCGCGGTGCGACAGCTCTCGCGGAGAGG 1391
Qy 869 GTATAGTACTGCTGGTCCGCGCGGCGGAACAGCGGCGCGCAACCTACAGCGTGGCTAG 928
Db 1392 GGTCTCTGTTCCGCTACGCGCGCGCGCGGAGAGGGCGCG-----GAGTCGATCGGTTGCG 1445
Qy 929 CGCGCGCGGAGGAGGCTATACCGTCTGGTTCAGTTGACAGCAAGCAACAGCTGGCA 988
Db 1446 CGGCGAGCGGAGCGCGCTCAGCGTCCGCGCGCTCGAGCAAGAGAGCTGGCG 1505
Qy 989 GCTTCTCCAGCAGGGAGCGAG---CGCGGAGGAGGCTCAACCGGGAAGTCTGGCGCC 1045
Db 1506 ACTTCTCTCCAGCGCGCGCGCGCTCGGCGAGCGCGGCTCAAGCGGAGGTCACCGGTC 1565
Qy 1046 CGCGGTCTGACATAGCCCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1100
Db 1566 CGCGGTCTGACATAGCGCGCGCTCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1625
Qy 1101 ---AAGCGACTACTACAGGAGGCTCTGGAACAGGAGGAGGAGGAGGAGGAGGAGGAGG 1156
Db 1626 AGGAGCGCGCGCTACATACATCTCGGCGAGCGGCTCGGCGAGGAGGAGGAGGAGGAGG 1685
Qy 1157 GCGTGGCGGCTCTCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1216
Db 1686 GCGCGCGCGCGCTCTGAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1745
Qy 1217 CCTCATGCGAGAGCGGAGATAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1276
Db 1746 CGCTACCGGCTCCAGGAGGCGG-----AAGTACCGCGCTCGAGGAGGAGG 1756
Qy 1277 CGGTAGGCTGAGCTCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1766
Db 1797 CGGCGGAGTCCAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
Qy 1337 CGGCTCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1386
Db 1857 TGAGCTCGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1906

RESULT 15

US-09-000-016-3

Sequence 3, Application US/09000016

Patent No. 6143541

GENERAL INFORMATION:

APPLICANT: Akita ARISAWA et al.

TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPIRIDINE DERIV

TITLE OF INVENTION: ITS EXPRESSION PRODUCT

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESS: Wenderoth, Lind & Ponack, L.L.P.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/000,016

FILING DATE: January 30, 1998

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2539 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: Streptomyces viridosporus

STRAIN: A-914

FEATURE:

NAME/KEY: CDS

LOCATION: 338...2535

IDENTIFICATION METHOD: E

US-09-000-016-3

Query Match 5.8% Score 190; DB 3; Length 2539;

Best Local Similarity 54.4% Pred. No. 1.4e-29;

Matches 582; Conservative 0; Mismatches 440; Indels 48; Gaps 8;

Qy 332 TGATAGACAGGGTACTTCTGTAACACACAGGGTCTCGGGCATAAAGTTTACACAGAGG 391
Db 870 TGTGGGAGCGCTCACCACAGGGGCGAGCGCGCTCGGGATCGCCAGCTCGGCTGG 929
Qy 392 ATTACAGGTTTCAGTTGACAGCGGCTCTCGGTCTCCAGATAGGGGCGGATACCGTCT 451
Db 930 AGGGGTTCGAGGCGCGGCTCGGACAGCTCGGACAGCTCGGCGGAGTCCGCGGCGG 989
Qy 452 GGAATCTCCCTGGGTACGAGGAGGCGTCTGCTGTTGCCATCTCGATACGGGTATAG 511
Db 990 GG---TCGCGCGGCTACAGCG3CAAGGGCGTGAAGATCGCGTCTTGACACCGGTTCG 1046
Qy 512 AGCGGAGCGCGGATCTGAAGGGCAAGGTCATAGGCTGTAGGAGCGCGGTCACCGGA 571
Db 1047 ACAGGAGGATCCGAGGAGTCAAGGGCGGGTGA---CGCGCTCCAGAGCTTACCGCGG 1103
Qy 572 GGTGACACCGCTTACGATGACAGGAGACAGGACCGGAGTTCGGGTATCGTTGCCGGA 631
Db 1104 CGCGCGGCGCGGCGACAGGTTGGGCGACCGGACCGAGCTCGATCGGCGGCGGA 1163
Qy 632 CGGGAGC---GTTAATCTCAGTACATAGGGTTCGCGCGGCGGCGGAGCTCGTGGCG 688
Db 1164 CGGGCGCGAGTCCAGGAGGAGTACAGGGGTGCGACCGCGCGGCGGATCTTCAAG 1223
Qy 689 TCAAGTTCTCGGTTCGCGAGGTTTCGGAAGGCTCTCACCATATCGCGGCTGTACT 748
Db 1224 GCAGGTCCTCGAGGAGTCTCGGTTTCGCGGAGGAGTCTCGGCGGATCTCGCGGAT 1283
Qy 749 GGTGCTCCAGACAGGAGTACGAGTACGGTCAAGGTCATCAACCTCTCCCTCGGCTCT 808
Db 1284 GGG-----CGGCGGCGAGGCGCGGAGGCTGTGTCAGCTAGGCTTGGCGGGA 1331
Qy 809 CCAGAGGTCGAGGAGGAGGAGTCTCCCTCAGTTCAGGCGCTCAACAGCGCTGGAGCGG 868
Db 1332 TGGACACACCGGAGAGGAGTCCGCGGCTGGAGGCGGCGTCCAGAGCTCTCGCGGAGAG 1391
Qy 869 GTATAGTACTGCTGCTCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 928

Db 1392 GCGTCTCTGTCGATCGCGCGCGCAACAGAGGCGCG-----GAGTCGATCGGTTGCG 1445
Qy 929 CCGCGCGCGCGAGCAAGGTGATACGGTGGGTGCAAGTTGACAGCAACGACACATCGGCA 988
Db 1446 CCGGCAAGCGCGACCGCGCGCTACCGGTGGGCGCGGTGCGAGACAGACAGCTCGCGG 1505
Qy 989 GCTTCTCCAGCAGGAGCGACGAC---CGCGGAGCGGAGCGCTCAAGCGGAGGTGTCGCGCG 1345
Db 1506 ACTTCTCTCTCCACCGCGCGCGCGCTCGGCGAGCGGCGCATCAAGCGGAGGTGTCGCGCTC 1565
Qy 1046 CCGGCGGTTGACATGACCGCGCGCGCGCGAGCGGAGCGATGCGCACCGCGAT----- 1100
Db 1565 CCGGCGGTGACATGACCG 1625
Qy 1101 ----AAACGACTACTACACCAAGCGCTCTGGAAACAGATGCGCACCGCGCGCGCGCG 1156
Db 1626 AGGCGCGCGCGCGCTACATGACCATCTCGGCGACGTGCGAGACCGCGCGCGCGCGCGCG 1685
Qy 1157 GCGTTGGCGGCTCATCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1216
Db 1686 GCGCGCGCGCGCTCTGAAGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1745
Qy 1217 CCTCATCGAGACCG 1276
Db 1746 GGTTCACCGCGCTCCACCG 1796
Qy 1277 CGGCTAGGCTGACAGCTCTACAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1336
Db 1797 CGGCGCGGATCCAGCG 1856
Qy 1337 CCGGCTCCGTCG 1386
Db 1857 TGAGCTTCGCGGTCCAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906

Search completed: November 2, 2002, 11:26:38
Job time : 72.5 secs

121 GGACTGCTGACCCGCGACTGTTCAAGAAAGTCTCCAGAGGATGAATGGAACGAGAGTGC 180
181 GACACCGTCATATGTTTCGGGAGCTACGGAGACAGGAGGAGCGGTTAAAGTACTAGG 240
181 GACACCGTCATATGTTTCGGGAGCTACGGAGACAGGAGGAGCGGTTAAAGTACTAGG 240
241 CTCATGGGCGCCAGGTCAGTCTCCTACAGATATCCCTGCTGTCGCGTTAAATA 300
241 CTCATGGGCGCCAGGTCAGTCTCCTACAGATATCCCTGCTGTCGCGTTAAATA 300
301 AAGGCGAGGAGCTTCTGCTGATCGCGGCGATGATAGACGCGGTTACTTCGGTAACACA 360
301 AAGGCGAGGAGCTTCTGCTGATCGCGGCGATGATAGACGCGGTTACTTCGGTAACACA 360
361 AGGGTCTGGGCTAAAGTTTCATACAGAGGATTAAGGTTTCAGGTCAGGAGCGGCT 420
361 AGGGTCTGGGCTAAAGTTTCATACAGAGGATTAAGGTTTCAGGTCAGGAGCGGCT 420
421 TCGGTCCTCCAGATAGGCGCGATACCGTCTGGGACTCCCTCGGCTACACGAGAGCGGT 480
421 TCGGTCCTCCAGATAGGCGCGATACCGTCTGGGACTCCCTCGGCTACACGAGAGCGGT 480
481 GTGGTGTTCGCTGCTGATACGGGTATAGCGGAACACCGGCTGATCTGAAGGCGAAG 540
481 GTGGTGTTCGCTGCTGATACGGGTATAGCGGAACACCGGCTGATCTGAAGGCGAAG 540
541 GTCATAGGCTGTACGCGCGCTACAGCGAGGTGACCGGCTACCGCTACGATGACGAGGAC 600
541 GTCATAGGCTGTACGCGCGCTACAGCGAGGTGACCGGCTACCGCTACGATGACGAGGAC 600
601 GGAACCCAGTTCGCGGCTATCGTTCGCGGACCGGCAUGGTTAACTCCGACTACATAGGC 660
601 GGAACCCAGTTCGCGGCTATCGTTCGCGGACCGGCAUGGTTAACTCCGACTACATAGGC 660
661 GTCCGCGCGCGGAGCTCGTTCGCGGCTCAAGGTTCTCGGTGCGGAGCGGTTTCGGGAGC 720
661 GTCCGCGCGCGGAGCTCGTTCGCGGCTCAAGGTTCTCGGTGCGGAGCGGTTTCGGGAGC 720
721 GTCTCCACCATATCCCGGCTGTGACTGGTTCGCGGAGCAAGGAGGAGGAGGAGGATA 780
721 GTCTCCACCATATCCCGGCTGTGACTGGTTCGCGGAGCAAGGAGGAGGAGGAGGATA 780
781 AGGTCATCAACCTCTCCCTCGGCTCCCGAGAGCTCCGAGGAGGAGGAGGAGGAGGATA 840
781 AGGTCATCAACCTCTCCCTCGGCTCCCGAGAGCTCCGAGGAGGAGGAGGAGGAGGATA 840
841 CAGGCGCTCAACACCGCTGGGAGCGGCTATAGTCTGCTGCGCGCGGCGGAGGAGGATA 900
841 CAGGCGCTCAACACCGCTGGGAGCGGCTATAGTCTGCTGCGCGCGGCGGAGGAGGATA 900
901 GGGCGGAGCAGCTACACCTCGGCTACCGGCTACCGGCGCGGCGGAGGAGGAGGAGGATA 960
901 GGGCGGAGCAGCTACACCTCGGCTACCGGCTACCGGCGCGGCGGAGGAGGAGGAGGATA 960
961 GCGGTTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020
961 GCGGTTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020
1021 AGGTCATAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1080
1021 AGGTCATAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1080
1081 ACCAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1140
1081 ACCAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1140
1141 ACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1200
1141 ACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1200
1201 GACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1260
1201 GACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1260

1261 GACATCGCTACGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1320
1261 GACATCGCTACGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1320
1321 GCCAAGCTCACCTTACACCGGCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1380
1321 GCCAAGCTCACCTTACACCGGCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1380
1381 GTCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1440
1381 GTCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1440
1441 GACCTTACCTTACGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1500
1441 GACCTTACCTTACGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1500
1501 GCGTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1560
1501 GCGTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1560
1561 AGCTACAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1620
1561 AGCTACAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1620
1621 TCGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1680
1621 TCGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1680
1681 CAGACCTTACCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1740
1681 CAGACCTTACCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1740
1741 GTCACAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1800
1741 GTCACAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1800
1801 CTCGAGCTTACCTTACGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1860
1801 CTCGAGCTTACCTTACGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1860
1861 AGCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1920
1861 AGCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1920
1921 GCGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1977
1921 GCGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATA 1977

RESULT 2
US-10-090-624-15
Sequence 15, Application US/100-0624
Patent No. US2002013235A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hiroyuki
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOTO, Tomoko
APPLICANT: ASADA, Kiyoko
APPLICANT: KATO, Ikunoshige
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERHYPERMOTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-01-06
PRIOR APPLICATION NUMBER: 06/45,472
PRIOR FILING DATE: 1998-12-05
PRIOR APPLICATION NUMBER: 31969/1997
PRIOR FILING DATE: 1997-06-23
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 15
LENGTH: 1962
TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-15

Query Match 37.9% Score 750.2; DB 12; Length 1962;
Best Local Similarity 64.7% Pred. No. 1.6e-163;
Matches 1219; Conservative 0; Mismatches 618; Indels 48; Gaps 5;

QY 110 AGAAGAACTACGGACTGCTGACCCGGGACTGTTCAAGAAAGTCCAGAGGATGAAGTGA-169
DB 107 AGAAGAACTATGCTGCTGCTAAGCCGAGGACTGTTCAAGAAAGTTCAGAAATTTGAATCTCA-166
QY 170 ACCAGAACTGGGACCGCTCATAAATTTGGGAGCTACGGAGACAGGACAGGCGGTTA-229
DB 167 ACGAGGAATACGACAGTAAATTTGTTGAAACCAATAGGAAAGAAATTTGCAATTA-226
QY 230 AGGTACTGAGGCTCATGGGCGCCAGGTCAAGTACTCCTACAAGATAATCCCTGCTGTCG-289
DB 227 GAGTCTTGAGTTAATGGGTGCMAAGTTAGGTATGTACCATATTATACCCGCAATAG-286
QY 290 CGGTAAATTAAGGCGAGGACCTTCTGCTGATCGGGGCGATGATAGACAGGCTTACT-349
DB 287 CTGCGGATCTTAAGGTTAGAGACTTACTAGTCACTCAGGTTTAACAG-----334
QY 350 TCGGTAAACAAAGGCTCTGGGCGATAAAGTTTCAACAGGAGGATTAACAGGTTTCAAGTT-408
DB 335 GGGTTAAGCTAAGCTTTACAGGTGTTAGGTTTATCCAGGAAGACTACAAAGTTACAGTTT-394
QY 408 -----GAGCAGCCCACTTCCGCTCCAGATAGGAGTATGGCAACTTACGTTTGG-454
DB 395 CAGCAGAAATTAAGAGGACTGGATGAGTCTGCAAGTATGGCAACTTACGTTTGG-454
QY 455 ACTCCCTCGGCTACGAGGAGGCTGTGGTGTGCTGATCGGTGATAGAGGCTATAGAGG-514
DB 455 ACT---TGGGATATGATGGTTCTGGAATCACAATAGGAATAATTGACACTGGAATTTGAG-512
QY 515 CGACACCCCGATCTGAGGGGCAAGGCTATAGGCTGTAGGCGGCGCTCAAGCGCAGGT-574
DB 512 CTCTCTATCCAGATCTCCAGGAAAGTAAATTTGGGTGGGTAGATTTTGTCAATGTTAGGA-571
QY 575 CGACCCCTAGTATGACGAGGACAGGAAACCCAGCTTGGGGGTATGTTTGGCGGAACCG-634
DB 572 GTTATCCATACGATGACATGGAATGGAATCTATGAGTTCATATAGCAGCTGTGACTG-632
QY 635 CGACGG---TTAACTCCAGTACATAGGCGTCCGCGCGGCGGCAAGCTCGTGGGCTCA-691
DB 632 GAGCAGCAAGTAAATGGCAAGTACAGGAAATGGCTCCAGGAGCTTAAGCTGGCGGGAATTA-691
QY 692 AGGTTCTCGGTGCGGACCGGTTCCGGAGCGCTCCACCATCATCGGGGTGTTGACTGG-751
DB 692 AGGTTCTAGTGGCGGATGGTTCTGGAAGCATATCTACTATTAATTAAGGGAGTTGAGTGG-751
QY 752 TCGTCCAGCAAGCAAGTACGGATAGGGTATAGGCTCATCAAGCTCTCCCTGGGCTCTCC-811
DB 752 CCGTTGATTAACAAAGATAAGTACGGAATTAAGTCTATTAATCTTCTCTGTTCAAGCC-811
QY 812 AGAGTCCGACGGAACCGACTCCCTCAGTCAAGGCGCTCAACAGCGCTGGAGCGGGTGA-871
DB 812 AGAGTCCAGATGGTACTGACGCTCTAAGTCAAGCTGTTAATGCAAGCTGGGATGCTGGAT-871
QY 872 TACTAGTCTGGCTCGCGCGGCAAGCGGCGGCAACACCTACACCGTGGGCTCACCG-931
DB 872 TACTGTTGTTGCTGCTGGAGCAAGTGGACCTTAACAAAGTATACAACTGTTCTCCAG-931
QY 932 CCGCGCGGAGCAAGGTATTAACGCTGGGTGAGTTGACGAGCAAGCAACATCGGCGCT-991
DB 932 CAGCTGCAAGCAAGTATTAACAGTTGGAGCGGTTGACAAGTATGATGTTATAACAGCT-991
QY 992 TCTCAGCAGGGGACCGGCGGAGGCTCAAGCGGAGTGTGCGGCCCGCGGCG-1051
DB 992 TCTCAAGCAGAGGCGCAACTGACAGCGGAGGCTTAAGCCCTGAGGTTGTTGCTCCAGGA-1051

QY 1052 TTGACATCATAGCCCGCGCGGCAAGCAAGCATGCGCACCCCGCATTAACGACTACT-1111
DB 1052 ACTGGATANTTGTCTGCCAGAGCAAGTGGAACTAGCATGGGTCAACCAATTAATGACTATT-1111
QY 1112 ACACCAAGGCTCTGGAAACCAAGCATGGCCACCCGACAGTTTTCGGGGGTTGGCGGCTCA-1171
DB 1112 ACACAGCAGCTCTCTGGGACATCAANTGGCAACTCTCAGGTAGCTGGTATTTGCAAGCCCTCT-1171
QY 1172 TCTCCAGGCGCCACCGGAGCTGGAGCCCGGCAAGCTGAAGACCGCTCATCGAGACCG-1231
DB 1172 TGCTCCAGCAGCAGCCGAGCTGGACTCCAGACAAAGTAAACAGAGCCCTCATAGAACTG-1231
QY 1232 CCGACATAGTCCGCCCCCAAGGAGATAGCGGACATCGCTACGGTTCGGGTAGGTTGAACG-1291
DB 1232 CTGATCTGTAAGCCAGATCAANTAGCCGATATAGCCTAGCGTTCAGGTAGGTTAATG-1291
QY 1292 TCTCAAGGCCATCAATTAAGCTAGCTAGCCCAAGCTACCTTCCACGGGCTCGCTGGCG-1351
DB 1292 CATACAGGCTATAAAGCTAGATAACTATGCAAGCTAGTGTTCACCTGGATATGTTGCCA-1351
QY 1352 CAAGGGAGCGCCACCGACACTTTCGAGCTGACGGGCGCCACCTTCTGTCAGCGCCACCC-1411
DB 1352 ACAAGGCGAGCAACATCACAGTGTGTTATTAGCGGAGCTTCGTTGCTGTAAGTCCACAT-1411
QY 1412 TCTACTGGCAGCAGCGGTTCAGGAGCATCGACCTCTACTCTAGGACCCCAAGCGGAACG-1471
DB 1412 TATACTGGGACAAATGCGCAATAGCGACTTGTATCTTACCTCTACGATCCCAATGGAAC-1471
QY 1472 AGGTGACTACTCTTACAAAGCTTACTAGGCTTCGAGAGGTGGGTACTACAAACCGGA-1531
DB 1472 AGGTGACTACTCTTACAAAGCTTACTAGGCTTCGAGAGGTGGGTACTACAAACCGGA-1531
QY 1532 CCGCGGAACTGAGGCTTCTAGGCTGCTAGCTTACAGGCGCGCGGCAACTTACAGGCTG-1591
DB 1532 CTGATGAAACATGGGAAATTAAGGTTGTAAGCTTACAGCGGAAAGTCAAACTATCAAGTAG-1591
QY 1592 ACTGCTCAGCAGCGGAGTCTCAGCGAGTTCGCGGCGGCAACCGCAATCCAAACCCCA-1651
DB 1592 ATGCTGTAAGTGTGTTCCCTTTCAGCGCTGGAG-----TTCACCAT-1636
QY 1652 ACCGAAACCAACCCGAGCTACCGACACCGACCTTACCGGTTTACCGGTTAAGGACTACT-1711
DB 1637 CTCACCAACCAAGCAAAATTAAGCAGCAAGAGCTTCCAAAGCATCCGATCACTACTACT-1696
QY 1712 GGACACAGCAGCAGCTTCAACATGAAGCTCAACAGCGGTGCCACCAAGATTAACCGGTG-1771
DB 1697 ATGACAGGAGCGAGCTTCAACATGAAGCTTAACTTGGGGCTACAAAGATTAAGTGGAG-1756
QY 1772 ACTGACCTTCGATATCTTCTACAAAGCTTCGAGCTCTACTCTTACGACCCCAAGCGCA-1831
DB 1757 ACCTAGTGTTCACACAAGTTACCATGATCTTGGAGCTTACGCTTACGATCCTAACCGA-1816
QY 1832 ACCTGCTTACAGGCTTCAAGCTTCGAGCAAGCTTACGAGCAGCTGAGTACGCCAACCCCG-1891
DB 1817 AGCTGTGATAGTTCGAGAGTCCCAAGAGCTACGAAACCGCTAGAAATTAATTAACCCCG-1876
QY 1892 CCGCGGAGCACTGGAGGTTCTCTCTTACCGCTTACAGCACTTACGGCTGGCGGAGTACC-1951
DB 1877 CCGCAGGAACTGGGACTTCTCTAGTATATCTCTTACGCTTACGCTTGGGCTTACTACG-1936
QY 1952 AGCTCAAGCGCTCTCTTACTACGG-1976
DB 1937 AGCTCAAGCGCTTAAGCTTATTAATG-1961

RESULT 3
US-10-090-624-2
Sequence 2, Application US/0090624
Patent No. US2002013235A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hiroyu
APPLICANT: MORISHITA, MAS
APPLICANT: SHIMOJO, TOMOAKI

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 20241
LENGTH: 1635
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006547.9
OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 17
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-20241

Query Match
Best Local Similarity 5.74; Score 111.8; DB 10; Length 1635;
Matches 514; Conservative 0; Mismatches 627; Indels 6; Gaps 2;
786 CATCAACCTCTCCCTCGGCTCTCCGAGAGTCCGACGAGACCGACGCTCTCGTCAAGC 845
|||||
388 CATCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 447
846 CGTCAACAGCGCTGGAGCGCGGTATAGTCTGCGTGGCGCGGCGGACGAGCGGCG 905
|||||
448 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 507
906 GAACACTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 965
|||||
508 TACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 567
966 TGACACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1025
|||||
568 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 627
1026 CAGCGCGGAGTCTGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1085
628 CACTGCGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1146
1086 CATTGCGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1205
688 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1265
1146 GCACCTTTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1325
748 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1405
1206 GTGGAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1465
808 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1525
1266 CCGCTACGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1385
868 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1445
1326 GCTCACTTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1505
928 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 987

QY 1386 CGCGCCACCTTCGTTGACCGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
DB 988 CACCATCACCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1043
QY 1446 CTACCTTACGACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1505
DB 1044 CCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1103
QY 1506 CGAAGAGTCTGCT 1565
DB 1104 CCACCATCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1161
QY 1566 CAAGGCGCGGAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1625
DB 1162 CACTTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1221
QY 1526 CGCGCGGCAACCGGANTCCCAACCGGCAACCGGCAACCGGCAACCGGCAACCGG 1585
DB 1222 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1281
QY 1686 CTTTACCGGTTTCTGTTTAAAGCTACTTGGGACACGAGGACCTTTCACCATGAC 1745
DB 1282 CATCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1341
QY 1746 CAGCGGTCGACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1805
DB 1342 CACTTCCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 1401
QY 1806 CTTTACTCTTACGACCGGCAACCGGCAACCGGCAACCGGCAACCGGCAACCGG 1865
DB 1402 TTCCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1461
QY 1866 CGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1925
DB 1462 CACCACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1521
QY 1926 CAGCACC 1932
DB 1522 CACCACC 1528

RESULT 5

US-09-864-761-3471
Sequence 3471, Application US/09864761
Patent No. US20030048763A1
GENERAL INFORMATION:
APPLICANT: Penn Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David F.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aomicca-X-
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 58/207,456
PRIOR FILING DATE: 2000-05-13
PRIOR APPLICATION NUMBER: US 59/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669

SEQ ID NO 3471

TYPE: DNA

**FEATURE:
OTHER INFORMATION**

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL

OTHER INFORMATION:	EXPRESSED IN HEART,	SIGNAL -
		9
		13

OTHER INFORMATION: EXPRESSED IN BI474, SIGNAL - II
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL -

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL -

S-09-864-761-3471

Query Match	5.7%	score	111.8;	DB 10;
				Le

Matches 514; Conservative 0; Miscellaneous 021; En

[illegible]

Q

1000

...
GAG GACGGGCGCTCAGCCGCCGAGCAAGGTCTATAG

662 TACCACCACCACCACCACCATCACTACCACCACCACCACCAT

966 TGACAGCAACGACAACATCGCCAGCTTCTCCAGCAGGGACCGACCG

722 CACCACCACCACCATCACTACCACCACCACCATCACATACCA

1026 CAAGCCGGAAGTCGTGCCTCCCGGTACATCATCA

..... 297

100

CAGGACCTT

[illegible]

1706 GGTGAAGACCGCCCTCATCGAGACCGCCGACATAGTCGCCCCCAAGG

962 CACCACCACCATCACTACCAACCACCACTCACTA

963 CACCACCACTACCAACCACCAACCATCTACTACCACCACCACCAC

RESULT 6

; sequence 4, Application 06/20069431A1
Patent No. US20020069431A1

APPLICANT: Broadway, Roxanne K.

1. TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITINOSIDASE AND THEIR EFFECTS ON DYING GROWTH AND DEVELOPMENT

; FILE REFERENCE: 1960373091
; TELETYPE TRANSMISSION NUMBER: 88-09748-033

CURRENT FILING DATE: 2008-07-21
PRIOR APPLICATION NUMBER: 02/0772,003

NUMBER OF SEQ ID NOS: 8

; SEQ ID NO 4

TYPE: DNA

4-20-60/48-05

100

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AL078472.1

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21

US-09-864-751-2513

Query Match 4.9%; Score 97.8; DB 10; Length 1403;
Best Local Similarity 46.0%; Pred. No. 9.7e-14;
Matches 452; Conservative 0; Mismatches 517; Indels 14; Gaps 3;

QY 885 CCGCCGCGGCAACAGCGGCGGCAACCTTACACGCTCGGCTCAGCGCGCGCGCGGAGCA 944
DB 1371 CACCAACAGTACCAACACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1312
QY 945 GGTGATACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1004
DB 1311 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1252
QY 1005 ACCGACCGCGGAGCGGAGGCTCAAGCGGGAAGTCTGCGCGCGGCGGCTGACATAGC 1064
DB 1251 CACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1192
QY 1065 CCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1124
DB 1191 CATGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1132
QY 1125 TGGAAACAGCATGGCCACCGCGGAGCTTTCGCGGCGGCTGATCTCTCAGCGGCA 1184
DB 1131 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1072
QY 1185 CCGGAGCTGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1244
DB 1071 CACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1012
QY 1245 CCGCAAGGAGATAGCGGAGATAGCGGAGATAGCGGAGATAGCGGAGATAGCGGAG 1304
DB 1011 CACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 952
QY 1305 CAAGTACGAGGAGTACGCGGAGGAGTACGCGGAGGAGTACGCGGAGGAGGAGGAG 1364
DB 951 CATCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 892
QY 1365 CACCCACACCTTGGAGGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1424
DB 891 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 833
QY 1425 GGGCTGAGGCGACATGACCTTACCTTACGAGCGGCGGCGGCGGCGGCGGCGGCG 1484
DB 832 -----TCACCATCAGCACTACCACTACCACTACCACTACCACTACCACTACCA 782
QY 1485 CTACACCGCTTACGCGTTCGAGAGGTCGAGAGGTCGAGAGGTCGAGAGGTCGAG 1544
DB 781 CCAACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 724
QY 1545 GAGGTCAGGTCGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAG 1604
DB 723 TACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 664
QY 1605 CCGGAGCTCAGCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1664
DB 663 CACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 604
QY 1665 CCGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1724

DB 603 CACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 544
QY 1725 CACCTTACCACTGAGCGTCAACAGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCG 1784
DB 543 TACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 484
QY 1785 TACTTCTCTACA--AGACCTTACGCTTACGCTTACGCTTACGCTTACGCTTACG 1842
DB 483 CACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 424
QY 1843 AGGTCCACGCTGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1865
DB 423 CACACTGCACTGATCCACCACTA 401

RESULT 9

US-09-790-399-7

Sequence 7, Application US/09790399

Patent No. US20020038000A1

GENERAL INFORMATION:

APPLICANT: Gold, Larry

APPLICANT: Tuerk, Craig

APPLICANT: Fridnow, David

APPLICANT: Smith, Jonathan D.

TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation

FILE REFERENCE: NEX02/CI-GON2

CURRENT APPLICATION NUMBER: US/09790,399

CURRENT FILING DATE: 2001-02-22

PRIOR APPLICATION NUMBER: 09/197,649

PRIOR FILING DATE: 1998-11-21

PRIOR APPLICATION NUMBER: 07/829,451

PRIOR FILING DATE: 1992-01-31

PRIOR APPLICATION NUMBER: 07/739,055

PRIOR FILING DATE: 1991-08-01

PRIOR APPLICATION NUMBER: 47/581,958

PRIOR FILING DATE: 1990-08-02

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO: 7

LENGTH: 390

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Sequence

OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed

OTHER INFORMATION: fragments having NcoI restriction sites.

US-09-790-399-7

Query Match

Best Local Similarity 4.9%; Score 84.8; DB 10; Length 390;

Matches 188; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 1247 CCAAGGAGATAGCGGAGATAGCGGAGATAGCGGAGATAGCGGAGATAGCGGAG 1306
DB 4 CCATGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 63
QY 1307 AGTACGACGACTACCGCAACGCTACCTTACCTTACCTTACCTTACCTTACCTTAC 1366
DB 64 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123
QY 1367 CCACACCTTTCAGCTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1426
DB 124 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183
QY 1427 GCTCGAGGACATCGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 1486
DB 184 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 243
QY 1487 ACACCGCTTACTAGCGTTCGAGAGGTCGAGAGGTCGAGAGGTCGAGAGGTCGAG 1546
DB 244 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 303
QY 1547 CGGTCAAGTCTGTCAGCTTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1606

Db 5860 ACACCCATCACCAACCACTACGCTGACCCCAACCCCAACCCACCCGCGCACAGACC 5819
Qy 1529 CGACCGCGGAACTGGAGGTCAGGTCTAGCTACAGGGGGGGGAACTACCAAGG 1588
Db 5920 CCAACGACGACACCACTACCAACCACTACGCTGACCCCAACCCCAACACCCAGCC 5979
Qy 1589 TCGAGCTGCTGACGACCGGAGGCTCAGGCTCGCGGGCGGCGCAACCCGAATCCAAAGC 1648
Db 5980 ACACAGACCCCAACGACGACACCCATCAGGCTGACCCCAACCCCAACCCCAACCA 6039
Qy 1649 CCAACCCGAAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 1708
Db 6040 CCAACCCGAAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 6099
Qy 1709 ACTGGGACACGACGACACCTTCAACGCTCAACGCTGACGCTGACGCTGACGCTGAC 1768
Db 6100 ACCCAACACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 6159
Qy 1769 GTGACCTGACCTTGGTACTTCTTACAGGACCTTCTTACGCTTCTTACGCTTCTT 1828
Db 6160 GTGACCTGACCTTGGTACTTCTTACAGGACCTTCTTACGCTTCTTACGCTTCTT 6219
Qy 1829 GCACCTGCTTGCAGGCTTCTTACAGGACCTTCTTACGCTTCTTACGCTTCTT 1888
Db 6220 ACCACTAGGCTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 6279

RESULT 12

US-09-833-263-1058
Sequence 1058, Application US/09833263
Patent No. US20020110547A1

GENERAL INFORMATION:

APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C12
CURRENT APPLICATION NUMBER: US/09/833,263
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 1058
LENGTH: 15720
TYPE: DNA
ORGANISM: Homo sapiens
US-09-833-263-1058

Query Match
Best Local Similarity 4.2%; Score 83.2; DB 10; Length 15720;
Matches 277; Conservative 0; Mismatches 323; Indels 0; Gaps 0;
Qy 1288 AGCTCTACAGGCTCATAGTACGAGGACCTACGCTGACCTTACCTTACGACCCCAACCCGA 1468
Db 5880 ATGACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1529
Qy 1349 CCGACAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1588
Db 5740 ACTACGCTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 5979
Qy 1409 CCCTCTACTTGGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1468
Db 5800 ACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5979
Qy 1469 ACAGGCTGCTACTCTTACACGCTTCTTACAGGCTTCTTACAGGCTTCTTACAGGCT 1529
Db 5860 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5913
Qy 1529 CCAACCCGAAACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1588
Db 5920 CCAACCCGAAACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5979

Qy 1589 TCGACGTGCTGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1648
Db 5980 ACACGACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 6039
Qy 1649 CCAACCCGAAACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1708
Db 6040 CCAACCCGAAACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6099
Qy 1709 ACTGGGACACGACGACACCTTCAACGCTCAACGCTGACGCTGACGCTGACGCTGAC 1768
Db 6100 ACCCAACACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 6159
Qy 1769 GTGACCTGACCTTGGTACTTCTTACAGGACCTTCTTACGCTTCTTACGCTTCTT 1828
Db 6160 GTGACCTGACCTTGGTACTTCTTACAGGACCTTCTTACGCTTCTTACGCTTCTT 6219
Qy 1829 GCACCTGCTTGCAGGCTTCTTACAGGACCTTCTTACGCTTCTTACGCTTCTT 1888
Db 6220 ACCACTAGGCTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 6279

RESULT 13

US-09-974-300-1934
Sequence 1934, Application 95/09974300
Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berka, Randy M.
APPLICANT: Clausen, IB Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
TITLE OF INVENTION: Expression
FILE REFERENCE: 10085-500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 05/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1934
LENGTH: 1329
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Query Match
Best Local Similarity 4.1%; Score 80.6; DB 10; Length 1329;
Matches 406; Conservative 0; Mismatches 359; Indels 60; Gaps 5;
Qy 472 GGAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531
Db 436 GGCACAGGAGTGCACAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 492
Qy 532 AAGGCGAAGCTCATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
Db 493 GAAGGCGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
Qy 592 CAGGCGACGGAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 648
Db 553 ATGCGGCGACGACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
Qy 649 CAGTACATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 708
Db 613 CAGTACCGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
Qy 709 GGTTCGGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 768
Db 673 GGATCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 732
Qy 769 AGTACGGGATAGGGGCTCATCAACTCT- 810
Db 733 GAAATCTGATGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGG 792

[illegible]

RESULT 15
 US-09-748-033-6
 Sequence 6, Application US/09748033
 Patent NO. US20020063431A1
 GENERAL INFORMATION:
 APPLICANT: BROGDWY, KORANNE M.
 APPLICANT: GONGORY, CARMENZA K.
 TITLE OF INVENTION: EFFECT OF ENDOCHITININASE AND CHITOSIDASE AND THEIR
 TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
 FILE REFERENCE: 19603/1091
 CURRENT APPLICATION NUMBER: US/09/748,033
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/172,003
 PRIOR FILING DATE: 1999-12-23
 NUMBER OF SEQ ID NOS: 8
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 6
 LENGTH: 1107
 TYPE: DNA
 ORGANISM: Streptomyces albidoflavus
 US-09-748-033-6

	Query Match	4.1%	Score 80.4	DB 10	Length 1107	
	Rest Local Similarity	46.4%	Pred. No. 9e-10			
Matches	403	Conservative	0	Mismatches 456	Indels 9	Gaps
777	GATAAGGGTCATCAGCCCTCTCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCCCT	836				
yy						
81	GATCACCCACATCAACTACTCTTCGGCAACGTCACAGGGCGGCAGTGTGCACATCGTGA	140				
bb						
837	CAGTCAGCGCGTCAACAACGGCTGGGACGGCGGTAAGTACTCTCGTCGCGCGCGCGGAA	896				
yy						
141	CAGCTTCGCGCGCTACGACATGCGGTACACCGCGCGGAGTCGAGCGCGTCCGCGA	200				
bb						
897	CAGC-GGGCGGAACACATCACTGTGCGGTCTACCCGCGCGCGAG--CAAGTCATAAC	953				
yy						

```

811  CAGAGCTCCGACGGAAACGAGCTCCCTCAGTCAGGCGCTCAACAAGCGCTGGAGCGCGCTT 870
Db  793  TACGAGAAATGAGAAGAGATCCAGTCGTTAAAGCTGTTTCATGACGATGTGGAGCGCAGCG 852
QY  871  ATAGTAGCTGTGCGTGTCCCGCGGACACAGCGGCGCGGAACACCTTACACGCTCGGCTCACCC 930
Db  853  ATTGTTTATATGTGCGGCGACGCGGCACTCCGCTCTCTGATGCGCAACAGTTGCCAGCGCG 912
QY  931  GCCGCCCGGACGAAGGTCATACCGTGGGTGCGAGTTGACAGCA----- 973
Db  913  GGTGTACGACGCAAGATTATTACAGTCGGAGCCTTGGATGACAGGGATACAGTCAGCGCG 972
QY  974  ----ACGACAACATGCCAGCTTCTCCAGCAGGGGACCGCGGACGAGAGGCTCAAG 1029
Db  973  GAGGATGACATGTCGCTCTATTTCAGCAGAGGGCGGACAATCTATGGTCAAGTCAAA 1052
QY  1030  CGGGAAGTCGTGCCCGCGGGTGTGACATCATAGCCCGCGCGGACCGGGAACACGAGCATG 1039
Db  1033  CGGGAATTGCTGTGTACGCGGCACAAATATTACGTCGCTCTGTTACCGGAGATCTTTCTC 1032
QY  1090  GGCACCCGAGATAACGACTAC-----TACACCAAGGCGCTCTGGACACGAC 1134
Db  1093  GATAGCTGTCAAAATACAAACAGAGTCGGCACAAATATATGACTGTTCGGAACCTCG 1152
QY  1135  ATGGCCACCCCGCAGCTTTCTGGGGGTGGCGCTCATCTCTCCAGGCGCCACCGGAGCTGG 1194
Db  1153  ATGGCTAGCGCGATCTGGCGAGGAATGCGCGCACTTATCTCTCAGCAAGCGCGCGGCACA 1212
QY  1195  ACCCGGACAAGGTGAAGACCGCCCTCATCGAGACCGCGGACATA 1239
Db  1213  GRACCTGATGAAGTCAACAGCTGCTATGAGCGCGGTACCGATTATA 1257

```

```

RESULT 14
US-09-974-300-5249/c
; sequence 5249, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berko, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5249
; LENGTH: 863
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(863)
; OTHER INFORMATION: n = A,T,C or G
; US-09-974-300-5249

```

[illegible]

Db 201 CACCTGGGACACCCCTCGCGGCAACTTCAACAGCTCCGCAAGCTCAAGGCAAGTA 260
QY 954 CTTGGTGCAGTTGACAGCAACAACTCGCCAGCTTCTCCAGCAGGGACCGACCG 1013
Db 261 CCCGACATCAAGTCTCTGCTTTCGGGGGTGACCTGGTCCGGGGCTTCACCGA 320
QY 1014 CGAGGAGCTCAACCCGAGTGTGTCGCCCCGGGGTTGAC---ATCATAGCTCCGG 1070
Db 321 CGCGTGAAGAACCCGGCGCTTCGCAAGTCTGCGACAGCTGGTGGAGACCGCG 380
QY 1071 CGCAGCGAACAGCATGGCACCCGATTAACAGACTACTACACAAAGGCTCTGGAAC 1130
Db 381 CTGGCCGAGGTTTCGACGGCTGACCTCGACTGGAGTACCGGACGCTCGGGCT 440
QY 1131 CAGCATGGCCACCCGACGTTTCGGGGCTTGGGGGCTCATCTCCAGGGCCACCGAG 1190
Db 441 CAGTTCGACAGCTCCGGTCCGGCGCGCTGGAAGAACATGGTCCAGGGGATGCGGCGCA 500
QY 1191 CTGACCCCGACAGGTGAAGACCGCCCTCATCGAGACCGCGGACATAGTCCGCCCA 1250
Db 501 GTTCGGCACCGACTTGTCCACCGCGCATCACCGCGAGCGAGCTCCGGGGCAAGCT 560
QY 1251 GGAGATAGCGGACATCGCTAGGTGGGGTGGGGTGAAGCTTACAAGGCCATCAAGT 1310
Db 561 CGAGCGCGGAGTACCGGGCGCGCCAGTACTTCGACTGGTACAACGTGATGAGTA 620
QY 1311 CGAGGACTAGCCCAAGCTCAGCTTACCGGC---TCCGTGCGGACAAAGGAGCGCCAC 1367
Db 621 CGACTTCTCGGGCTTGGGACAAAGACCGCGCGGACCGCGCCACTCGGCCCTGACATC 380
QY 1368 CCACACTTCGAGGTCAGGGCGGCGACTTCGTGACCGCCACCTCTACTGGGACAGGG 1427
Db 681 CTACAGGGCATCCCAAGCGGACTTCCACTCGGGCGCGCCATCGCCAAAGTCAAGGC 740
QY 1428 CTCAGGGAGATCGACTCTACCTTACGACCCCAACGGGAGAGGTTGACTACTCTTA 1487
Db 741 GAAGGGGTCCCGGGGAGCAAGCTCTGCTCGCATCGGCTTCTACGGCCCGGCTGGAC 800
QY 1488 CACCGCTACTACGGCTTCGAGAGGTGCGCTACTACAACCGACCGCGGMACTTGAC 1547
Db 801 CGGGTCACCCAGGACCGCGGGGGCGACCGCACCGCCCGCGCCACCGGCACTACGA 860
QY 1548 GGTCAAGGTCTCAGCTACAAGGGCGGGCGGAACTACAGGTGAGCTGCTCAGCGACGG 1607
Db 861 GGGGGGATCGAGGACTACAAGGTCTCAAGAGACACTGCGCGCGCACCGGACCGCTCG 920
QY 1608 GAGCTCAGCCAGTCCGGCGCGGCAAC 1635
Db 921 CGGCACCGGTACGGCAAGTCCGCGACG 948

Search completed: November 2, 2002, 13:08:29
Job time : 97.5 sec

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sv model

Run on: November 2, 2002, 08:10:53 : Search time 1873.5 seconds
(without alignments)
17090.193 Million Cell updates/sec

Title: US-09-841-553-2
Perfect score: 1977
Sequence: 1 ATGAGAGGTTAGTGCTGT.....AGGCCGCTGCTACTACCGG 1977

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estlin.*
 - 4: em_estmu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_estc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_est3.*
 - 12: gb_est4.*
 - 13: gb_est5.*
 - 14: gb_est6.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: gb_gss.*
 - 18: em_gss_hum.*
 - 19: em_gss_inv.*
 - 20: em_gss_pla.*
 - 21: em_gss_vrt.*
 - 22: em_gss_fun.*
 - 23: em_gss_mus.*
 - 24: em_gss_mus.*
 - 25: em_gss_other.*
 - 26: em_gss_pro.*
 - 27: em_gss_rtd.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	C	93.4	4.7	641	13	BM181884	BM181884 f51b11.y
2	C	86.4	4.4	681	17	CNS02EOD	AL193990 Tetraodon
3	C	86.2	4.2	500	17	B67199	B67199 CpG00158 Cp
4	C	81.6	4.1	440	10	B6537787	B6537787 WHE1755-1
5	C	78.8	4.0	1116	11	AV110435	AV110435 Zea mays
6	C	78.4	4.0	512	10	AV938606	AV938606 AV938606

7	77.2	3.9	482	14	BQ170719
8	76.4	3.9	750	17	AZ046831
9	76.2	3.9	705	14	BQ753023
10	75.8	3.8	422	13	B1098380
11	75.6	3.8	645	14	BQ753506
12	75.4	3.8	773	17	CNS01VTC
13	75	3.8	627	12	EG608098
14	73.6	3.7	559	10	AM4533235
15	73.6	3.7	607	14	BQ660503
16	73.6	3.7	649	13	B1641907
17	73.6	3.7	692	14	BQ766288
18	73.6	3.7	852	12	EG110300
19	73	3.7	585	10	BE470936
20	73	3.7	617	10	BE361069
21	72.8	3.7	627	13	BQ758334
22	72.8	3.7	645	9	AL389106
23	72	3.6	877	13	B1952167
24	71.6	3.6	443	13	BM039802
25	71.6	3.6	651	13	BQ744245
26	71.2	3.6	915	12	EG343471
27	71	3.6	534	10	BE495200
28	71	3.6	608	13	BQ733059
29	70.8	3.6	521	13	BM004465
30	70.4	3.6	1490	11	AV103602
31	70.2	3.6	820	13	B1956721
32	70	3.5	639	13	BM371945
33	69.4	3.5	525	13	BF727798
34	69.4	3.5	553	13	BQ366220
35	69.2	3.5	512	10	BE195101
36	69.2	3.5	2558	11	AV103647
37	68.8	3.5	317	13	BQ366180
38	68.6	3.5	733	10	AM587459
39	68.4	3.5	494	13	BQ360881
40	68.2	3.4	970	17	CNS0386V
41	67.8	3.4	690	12	BQ818224
42	67.8	3.4	701	14	BQ818521
43	67.4	3.4	677	13	BM569289
44	67.4	3.4	722	14	BQ766040
45	66.8	3.4	535	10	BE603252

ALIGNMENTS

RESULT 1
BM181884
LOCUS
DEFINITION
f51b11.y1 Sugarcane adult male Danio rerio cDNA clone 5412044 5', similar to contains element TAR1 repetitive element ;, mRNA
SEQUENCE
BM181884
VERSION
BM181884.1 GI:17512842
KEYWORDS
EST.
SOURCE
zebrafish
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 641)
AUTHORS
Clark, M., Johnson, S.L., Lebrach, H., Lee, R., Li, P., Marra, M., Eddy, S., Hillier, L., Lucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
TITLE
WashU Zebrafish EST Project 1998
JOURNAL
Unpublished (1998)
COMMENT
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1890
Fax: 314 286 1810
Email: zbrafish@wustl.wustl.edu

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA sequencing by: Washington University Genome Sequencing Center: Clone distribution information can be found through the I.M.A.C.E. Consortium/LINL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence atop: 527.

FEATURES

Location/Qualifiers
1. 641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SJD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site 1: DraIII (CACCATGTG); Site 2: DraIII (CAGGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor: (TGTGGCTACTGG), digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CAGGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTCTCTTAAGCTCG and 3' end primer CGACCTGGACGTCGACACA."

BASE COUNT 218 a 236 c 37 g 152 t
ORIGIN

Query Match 4.7%; Score 93.4; DB 13; Length 641;
Best Local Similarity 47.7%; Pred. No. 2.7e-09;
Matches 274; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

QY 1289 ACCTCTACAGGCCATCAGTACGAGCTACGACGCTCAGCAAGCTCACCTTCACCGGCTCCGTCG 1348
DB 65 ACCTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 124
QY 1349 CCGACAGGGAAGCGCCACCCACCTTCGAGCTCAGCGGGCGCCAGCTTCGTCACCGCCA 1408
DB 125 ACTGCTACTGCTACATCACTTCTCCACACACAACTTCTCCAGCTACATCACT 184
QY 1409 CCCTCTACTGGGACAGGCTCGAGCGAGCTGAGCTTACTCTAGACCCCAACCGGA 1468
DB 185 TCTCCACACACAACTTCTCCACACACAACTTCTCCAGCTACATCACTTCAACA 244
QY 1469 ACAGGTTGACTTCTCTACACCGCTTACTGCTTCCAGAGGTCGGCTACTACAC 1528
DB 245 ACACACAACTTCTCCAGCTACATCACTTCAACACAACTTCTCCAGCTACAC 304
QY 1529 CGACCGCGCACTCGAGGCTCAAGGTCGTGAGTACAGGGCGGGGCACTACCAAG 1588
DB 305 TCTCCAGCAACAACTCACTTCTCCAGCTACATCACTTCTCCACACAACTTCT 354
QY 1589 TCGAGCTGTCAGGAGGAGGCTCGAGCGAGCTGAGCTTCTAGACCCCAACCG 1648
DB 365 CCAGCTACATCACTTCTCCACACAACTTCTCCAGCTACATCACTTCTCCAGCA 424
QY 1645 CCAACCCGACCCACCCGACACCGACACCGACGCTTCCCGGTTCCGTTAGGACT 1708
DB 425 ACACAACTTCTCCAGCTACATCACTTCAACACAACTTCTCCAGCTACATCACT 484
QY 1709 ACTGGGACACGAGCACTTCCAGCTACACCTTCCAGCAACAACTCACTTCTCCA 1768
DB 485 ACTTCACACAACTTCTCCAGCTACACCTTCCAGCAACAACTCACTTCTCCA 544
QY 1769 GTGAGCTGAGCTTCTTCTTACAGGACCTGAGCTTCTTCTAGGCGCCCAAG 1828
DB 545 GCTACATCACTTCTTACAGGACCACTTCTTCCAGCTACATCACTTCTTCCAGCA 604
QY 1829 GCAACCTCTGTTGACAGGTCAGGTCGAGCAACAGC 1863

Db 605 AACATCTCCAGTACAACTTCTCCAGCAAC 639

RESULT 2

CNS02EOD/c 681 bp DNA linear GSS 13-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 and of clone
262H14 of library G from Tetraodon nigroviridis, genomic survey
sequence.

LOCUS

DEFINITION
ALL93990.1 GI:7837096
GSS: genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

/db_xref="taxon:77009"
 /clone_lib="K. Sato unpublished cDNA library, strain H602"
 /adult_heading="top three leaves"
 /tissue_type="top three leaves"
 /dev_stage="adult, heading stage"
 BASE COUNT 86 a 202 c 150 g 73 t 1 others
 ORIGIN

Query Match 4.0% Score 78.4; DB 10; Length 512;
 Best Local Similarity 49.2%; Pred. No. 3.3e-06;
 Matches 239; Conservative 0; Mismatches 241; Indels 6; Gaps 1;
 QY 1095 CCGATAACGACTACTACACCAAGCCCTCTGGACACAGATAGCCACCCCGCTTTC 1154
 DB 33 CCGAGGCGCTCCCGTTCAACATCATCTCCGGGACGTCTCATGTCTGCGCCGCTAG 92
 QY 1155 GCGCTTGCGCGCTCATCTCCAGGCCACCCGAGCTGGACCCGACAGGTGACAC 1214
 DB 93 CGGCTTCGCGCGCTGCTCCGGAGGCGACCCGAGTGGAGCCGCGGCTCAAGTC 152
 QY 1215 CCGCTCATCGAGACCCCGACATAGTCCGCCCAAGGATAGCGGATCGCTAGG 1274
 DB 153 GCGCTCATGACACCGCGTACA-----ACCTGGACACTCGGGAGATCATCAGGA 206
 QY 1275 TCGGGTAGGCTGAACCTCTAGAGGCCATCAAGTACGAGCTACGCCAGCTTCCTT 1334
 DB 207 CTGGCCACCGGAGGAGCTGACGCGCTTCTCCGGGCGGCGGACGTGGACCCCAA 266
 QY 1335 CACGGCTCGCTCCGACAAAGGAGCGGACACACACCTTCGACGTCAGCGGCGAC 1394
 DB 267 CAGCGCTCAACCGCGCTGGTGTAGGACGCGGACACCGCGACTCATCGCTTCT 326
 QY 1395 CTTCGTACCGGACCTCTACTTGGACAGGCTCGAGGACATCGACCTTACTCTTA 1154
 DB 327 CTGCGCGTGGCTACACTCCATCCAGATCGCGCTTCTCTAGGAGGCTCCGTCG 386
 QY 1455 CGACCCCAAGGAGGAGTGTACTTCTTACCGCTTACTAGCGCTTTCGAGAG 1114
 DB 387 CGACTGCTCGAAGAAGCGCGCGCTCCGCGGAGCTCAACTACCGCGCTTCGCGCG 445
 QY 1515 CGGCTACTACACCGGACCGCGGAACTGGACGCTCAAGCTGCTCAGTACAGGCGC 1574
 DB 447 CTTCGTGCTGACAAAGGACTCGCTCACCTACCCGCGGTGGTGGACAGCTCGG 506
 QY 1575 GCGGAA 1580
 DB 507 CCGGAA 512

RESULT 7
 BQ170719
 LOCUS
 DEFINITION
 accession BQ170719
 version BQ170719.1
 keywords
 source
 organism

482 bp mRNA linear EST 26-APR-2002
 BQ170719
 LOCUS
 DEFINITION
 accession BQ170719
 version BQ170719.1
 keywords
 source
 organism

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: candersnep.usda.gov

This EST was generated by sequencing from the 3' end of the clone.
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20.
 Seq primer: 77 primer.

FEATURES
 source
 Location/Qualifiers

1. 482
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone_lib="H61798_B12_D24"
 /tissue_type="Wheat pre-anthesis spike cDNA library"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOUR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site: EcoRI; Site 2: XhoI; Plants were grown in the
 greenhouse. Whole spike with awns trimmed, white, green
 and yellow anther were collected and total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the TJ clone lab (Choi, Close, Panton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."
 BASE COUNT 90 a 172 c 148 g 72 t

Query Match 3.9% Score 77.2; DB 14; Length 482;
 Best Local Similarity 55.4%; Pred. No. 5.7e-06;
 Matches 173; Conservative 0; Mismatches 131; Indels 6; Gaps 1;
 QY 979 AACATCCCAAGTCTTCACAGGGGACGACCGGG-----ACGGAGGCTCAGCCG 1032
 DB 141 ATCGTCCGCGCTTCTCGCTCCGCGGCGGACGCGCGCGGCTCTCAAGCC 200
 QY 1033 GAAGTGTGCGCGCGCGGCTTTCGACATAGCCCGCGCGGACGCGGAGCCAGCTGGC 1092
 DB 201 GACATATGCGCGCGCGGCTTCAACATCTCGCGCGTGGCGCTGATGCTATCAT 360
 QY 1093 ACCCGGATAAGAGCTACTTACCAAGGCTCTGGAAACGACGATGGCAGCCGCGCTT 1152
 DB 261 GGCACGAGTCTACAGCTTAAACGTTGCTTCGCGGCTGCTGCGCGCGCGCTCA 320
 QY 1153 TCGGCGCTTGGCGCGCTCTTCTCCAGGCGCGCGGCTGAGCCCGGACAGGTGAAG 1212
 DB 321 ACCGCTATCTGCGCGCTCTGTAAGAGGCGGCGCGGCTGCTGCGCGGTCAAG 380
 QY 1213 ACCGCTCTATCGAGACCGCGGACATAGTCCGCCCGCAAGACATAGCGGACATCGCTAC 1272
 DB 381 TCGGCTATCATGACCGCTCTCAGCAGCGCGGACAGCGGCGGCGGCTATCATGACGA 440
 QY 1273 GGTGCGGCTAGG 1284
 DB 441 GACACGGAAG 452

RESULT 8
 AZ046831
 LOCUS
 DEFINITION
 accession AZ046831
 version AZ046831.1
 keywords
 source
 organism

750 bp DNA linear GSS 16-MAR-2000
 AZ046831
 nbe0088017r CUGL Rice BAC Library (ECORI) Oryza sativa genomic
 clone OSJNB0088017r DNA sequence.
 AZ046831.1 GT 725640
 Oryza sativa.
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.


```

REFERENCE
AUTHORS      Wing,R.A. and Dean,R.A.
TITLE        A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL      Unpublished (1998)
COMMENT      Contact: Wing RA
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: rwing@clemson.edu
              Seq primer: GCTATTGAGTGACACTATAG
              Class: BAC ends
              High quality sequence stop: 339.
              Location/Qualifiers
                1. .750
                  /organism="Oryza sativa"
                  /strain="Japonica"
                  /cultivar="Nipponbare"
                  /db_xref="taxon:4530"
                  /clone="OSJNB0088017r"
                  /clone_lib="CUGI Rice BAC Library (ECORI)"
                  /tissue_type="Leaf"
                  /lab_host="E. coli DH108"
                  /note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
                  Rice is the most important food crop in the world. Half of
                  the world population, especially those inhabiting highly
                  populated areas of the humid tropics and subtropics, rely
                  on rice as their primary source of carbohydrate.
                  Monocotyledonous rice is a diploid plant (2n=24) with a
                  haploid genome equivalent of 431 Mbp (Arumuganathan and
                  Earle, 1991). The relatively small genome of rice, three
                  times larger than that of Arabidopsis, makes it suitable
                  for genomic studies. In order to facilitate positional
                  cloning, physical mapping and genome sequencing of rice,
                  we have constructed a BAC library from Oryza sativa.
                  Nipponbare variety using EcoRI as the cloning enzyme. The
                  library contains 55,296 clones with an average insert size
                  of 121 Kb providing approximately 15 haploid genome
                  equivalents. The deep coverage allows the isolation a
                  particular sequence with a probability of 99.9%. Three
                  high density filters, each containing 18,432 clones
                  (doubly spotted), represent the whole library for colony
                  screening and can be requested from the Clemson University
                  BAC/EST Resource Center (www.genome.clemson.edu).
                  BAC/EST Resource Center (www.genome.clemson.edu).
                128 a 254 g 110 t
                254 c 256 g 110 t
                2 others

BASE COUNT  128 a 254 g 110 t
ORIGIN

Query Match      3.9%; Score 76.4; DB 17; Length 750;
Best Local Similarity 54.4%; Pred. No. 9.4e-06;
Matches 204; Conservative 0; Mismatches 157; Indels 14; Gaps 2;

QY 1023 GCTCAAGCCGAGTGTGCGCCCGGCGTTGACATATAGCCCGCCCGCCGCGGAC 1074
DB 214 GCTCAAGCCGAGTGTGCGCCCGGCGTTGACATATAGCCCGCCCGCCGCGGAC 273
QY 1083 CAGCATGGCACCCTGATAACGA-----CTACTACACCAAGCCCTCTGGAAC 1130
DB 274 GGCACCGTCGGGCGTGGAGGGCGACGCCCGCGGCTCCGACTTCACGCTGCTCGGGC 333
QY 1131 CAGCATGGCACCCTGACGCTTTCGGGCGTTCGGCGCGTATCTCCAGCGCCACCCGAG 1190
DB 334 ATCAATGGCGTCCCGCAGCTGACGGCATCGCGCGCTCATCAGTTCGGCGCCACCCGC 353
QY 1191 CTGACCCCGGAGAGTGAAGACCGCCCTCATCGAGACCGCGACATAGTCGCCGCCAA 1250
DB 394 GTGGACCGCGGATGTTGCGGTCCCGCATCATGACGACCGCCGACGTCACGCGGCA 453
QY 1251 GGAGATAGGCGATCGCTACGCTGCGGTGGGTAGGTGA--ACGTCTACAAGGCCATCAAG 1308
DB 454 GGGCAAGCGGATCATGGACGGGACGCGAAGCGCGACGCGTACGCGATGGCGCGGG 513
QY 1309 TAGCAGACTAGCCCAAGCTCACCTTCACCGGCTCCGTCCGCCGACAGGAGCGCCACC 1366

REFERENCE
AUTHORS      BQ753023
TITLE        WHE4122_A06_A12S Wheat salt-stressed root cDNA library
JOURNAL      Unpublished (2002)
COMMENT      Contact: Olin Anderson
              US Department of Agriculture, Agriculture Research Service, Pacific
              West Area, Western Regional Research Center
              800 Buchanan Street, Albany, CA 94710, USA
              Tel: 5105595773
              Fax: 5105595818
              Email: oanderson@usda.gov
              Sequences have been trimmed to remove vector sequence and low
              quality sequence with phred score less than 20
              Seq primer: SK primer:
              Location/Qualifiers
                1. .705
                  /organism="Triticum aestivum"
                  /cultivar="Chinese Spring"
                  /db_xref="taxon:4565"
                  /clone="WHE4122_A06_A12"
                  /clone_lib="Wheat salt-stressed root cDNA library"
                  /tissue_type="Roots"
                  /dev_stage="Full tillering"
                  /lab_host="E. coli SOLR"
                  /note="Vector: Lambda Uni-ZAP XR, excised phagemid
                  paluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Hydroponic
                  plants grown to full tillering stage were treated with 150
                  mM NaCl for either 12 hours or 7 days. Root tissues of the
                  plants subjected to both types of treatment were collected
                  separately at University of California, Davis (E. Akhunov
                  and K. Deal in J. Dvorak's Lab). Total RNA was prepared
                  and equal amount of RNA was then pooled. PolyA RNA was
                  purified from the pooled RNA, a cDNA library was made, and
                  the cDNA clones were in vivo excised to give pBluescript
                  SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak)
                  at the University of California, Davis. Colony plating,
                  plasmid DNA preparations and DNA sequencing were performed
                  in the Q2 Anderson lab (all other authors)."
                157 a 280 g 201 g 67 t
                280 c 201 g 67 t

BASE COUNT  157 a 280 g 201 g 67 t
ORIGIN

Query Match      3.44; Score 76.2; DB 14; Length 705;
Best Local Similarity 50.7%; Pred. No. 1e-05;
Matches 265; Conservative 0; Mismatches 248; Indels 10; Gaps 3;

QY 1367 CCACACCTTGACGTCACGGCGCCACCTTCGTGACCGCCACCTCTACTGGACACGG 1426
DB 157 CCACCTTGTCGGCGCTCTCTCTGGCGACTCCATCTGTCGCCCGCGGACGCCGCC 216

```

QY 1427 GCTGAGGACATCGACCTTACCTCTACGACCCCAAGGAGGTTGACTACTCCT 1486
 DB 217 GCGCCGCAACGCCCGCGCTCTTCTCCGCTCCGAGGGCCCAACAAAGCGG 276
 QY 1487 ACACCGCTACTACGCTTCAGAAAGTGGCTACTACAAACCGACCGCGAACCTTGA 1546
 DB 277 GACCGGATTCGCCGCCGCCCAAGATGAAGCGGAGCTTCAAGGGCCCGGCTCATCGCT 336
 QY 1547 CGGTCAAGTCTGCTAGCTTACAAAGGGGCGCGAACTTACAGTCTGAGTCTGTCAGGAG 1506
 DB 337 CCGGCAAGTCTGCTTCCG---GCCACCAACCAAGGAGCGGCTTACAAACAGC 393
 QY 1607 GAGCCTCAGCAGTCCGGGGGCGGAGCCGCAATCGAATCGAAGCCCA-----CCGCAACCA 1662
 DB 394 TCAGCTCTCCCGCGCGCGGAGCCCAAGCCCTTACGCAAGCGGAGCGGCTTACGCT 453
 QY 1663 ACCCGACCAACGACCGACAGCTTTCACCGCTTCCGTTTACGACTACTGGGACACACG 1722
 DB 454 CCNAGCCCAACAGCAATCAACAGGCTTCGGGCTCAACAGATGGCTGCCGGCGC 513
 QY 1723 GACACCTTACCA---TGAAGTCAACAGCGGTGCCACCAAGATACCGCTGACCTGACC 1779
 DB 514 GGGGCTACAGCAACTTCAACGCGCGCAACGAGGGGTGAGAGCTACTTCAACAGTCC 573
 QY 1780 TTGATACTTCTCAACAGCACTCGACCTTCTACCTCTACGACCCCAACCGCAACCTGTT 1839
 DB 574 ATCGGAGCGCGCGGAGCAACAAACAGCAGCAGCAACACTTACGCGGCTTACGCGGC 633
 QY 1840 GACAGTCTGCTGACCAACAGCTTACGAGCAGCTCGAGTACG 1882
 DB 634 AAGAAGGCGCGCGGAGCGCAAGAGAGAGCAACGCAAGAGC 676

RESULT 10
 LOCUS BI098380 422 bp mRNA linear EST 26-JUN-2001
 DEFINITION IP1_3_E03.bl_A002 Immature pannicle 1 (IP1) sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BI098380
 VERSION BI098380.1 GI:14569948
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1. (bases 1 to 422)
 AUTHORS Klein, R.R., Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M. and Pratt
 L.H.
 TITLE An EST database from Sorghum: developing preanthesis pannicles
 JOURNAL Unpublished (2001)
 COMMENT Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mpratt@uga.edu
 Sequences have been trimmed to exclude polyA, vector and regions
 below Phred quality 16. The threshold for high quality sequence is
 20. Three-prime sequences, which are obtained with PolyTm1 or J7
 sequencing primer, are presented as the reverse complement.
 Seq primer: JEN REV
 High quality sequence stop: 419
 POLYA=NO. Location/Qualifiers
 1..422
 /organism="Sorghum bicolor"
 /cultivar="BTx623"
 /db_xref="taxon:4558"
 /clone_lib="Immature pannicle 1 (IP1)"
 /note="Organ: Developing preanthesis pannicles; Vector:
 pBluescript II SK(-) from Lambda zap II; Site_1: XhoI;
 POLYA=NO. Location/Qualifiers

FEATURES
 source
 1..645
 /organism="Sorghum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone_lib="Eban01_SQ004_H09"
 /note="Site: anther, yellow stage, no treatment, cv Optic,
 Eban01."
 /tissue_type="anther"
 /dev_stage="yellow stage"
 /lab_host="DH10B"
 /note="Vector: pSPOR1; Site_1: Sal I; Site_2: Not I;
 Non-normalized library, directionally cloned into pSPOR1.
 Derived from anthers dissected from developing flowers in
 glasshouse grown barley plants. Developed as part of the
 barley transcriptome resources of BBSRC/SEPRAD funded

Site 3: EcoRI. The library was made from poly-A RNA in the
 cloning vector lambda zap II. Clones to be sequenced were
 prepared by mass excision.

BASE COUNT 75 a 155 c 135 g 57 t

Query Match 3.9%; Score 75.8; DB 13; Length 422;
 Best Local Similarity 5.5%; Pred. No. 1.1e-05;
 Matches 152; Conservativ 0; Mismatches 127; Indels 0; Gaps 0;

QY 1018 GGAAGGCTCAAGCGGAGTCTGCTGCCCGCGGCTTACATACATACCCCGCGCCGAGC 1077
 DB 114 GCGCTGCTGAAGCGGAGTCTGCTGCCCGCGGCTTACATCTCTCGGCGTGGCGCTG 173
 QY 1078 GGAACAGCATGGGACATCGGATAAAGCACTACTACCAAGCGCTTCTTGAACACGATG 1137
 DB 174 GAGTGCGGCTGCGGGGCTCCGAGTCTCAAGAGCGTCAAGTCTTCTCGGCGACGTCATG 233
 QY 1138 GCGACCGCGACGCTTTGCGGCGCTTATCTCCAGCGCCACCGAGCTGGACC 1197
 DB 234 GCGACCGCGACATACCGCTATCTGCGCTGCTCAAGAGCGGCGACCGGAGTGGTCG 293
 QY 1198 CCGGACAAAGTGAAGACCGCGCTCATCGAGACCGCGGAGATACCTCGCCCGCAAGGAGATA 1257
 DB 294 CCGCGCGGATCAAGTCCGCCATCATGACCACTGCCAGCGCTCGACAGCGGCGGAC 353
 QY 1258 GCGGACATCCCTTACGCTGCGGCTAGGCTGAGGCTGAGCTCTAC 1296
 DB 354 CAGATCATGAGGAGGAGCAGCGGAGCGAGCTTCTAC 392

RESULT 11
 LOCUS BI0753506 645 bp mRNA linear EST 26-JUL-2002
 DEFINITION Eban01_SQ004_H02_R.anther, yellow stage, no treatment, cv Optic,
 Eban01 Hordeum vulgare cDNA clone Eban01_SQ004_H09 5', mRNA
 sequence.
 ACCESSION BI0753506
 VERSION BI0753506.1 GI:21961978
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 REFERENCE 1. (bases 1 to 645)
 AUTHORS Hedley, P., Liu, K., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
 Ramsay, L., Mackay, G., Marshall, D.F.M. and Waugh, R.
 TITLE Development of Barley Transcriptome Resources
 JOURNAL Unpublished (2002)
 COMMENT Contact: Waugh, R., Marshall, D.F.
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee DD2 5DA, Scotland, UK
 Tel: 00 44 1382 82731
 Fax: 00 44 1382 82746
 Email: est@scri.sari.ac.uk
 Location/Qualifiers
 1..645
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone_lib="Eban01_SQ004_H09"
 /note="Site: anther, yellow stage, no treatment, cv Optic,
 Eban01."
 /tissue_type="anther"
 /dev_stage="yellow stage"
 /lab_host="DH10B"
 /note="Vector: pSPOR1; Site_1: Sal I; Site_2: Not I;
 Non-normalized library, directionally cloned into pSPOR1.
 Derived from anthers dissected from developing flowers in
 glasshouse grown barley plants. Developed as part of the
 barley transcriptome resources of BBSRC/SEPRAD funded

```

BASE COUNT      124 a 202 c 233 g 86 t
ORIGIN
Query Match      3.8%; Score 75.6; DB 14; Length 645;
Best Local Similarity 55.2%; Pred. No. 1.3e-05;
Matches 169; Conservative 0; Mismatches 134; Indels 3; Gaps 1;

QY 1106 ACTATGACCAAGGCTCTGACACCATGCGACCGCCGACCTTCGGCGCTGGCG 1155
DB 14 ACAGTACAGGTTGCTCTCGGACGCTCAATGGCCAGCCGACGACGCTCGGCTGG 73
QY 1166 GCTCATCTCTCAGGCGCCACCGAGCTGGACCCCGGACAGGTGAAGACCGCTTCG 1225
DB 74 CACTGTTGAAGAGGCGACCCGAGCTGGTGGCGTGGCGGTCAGTCCGCCATCGA 133
QY 1225 AGACCGCGACATAGTCCGCCCAAGAGATAGGACATCGCTACGTCGGGTAGG 1285
DB 134 CCAGTTCAGACCGCCGACACGCGGGGACCCCATCATGAGGAGGACCGGAGG 193
QY 1286 TGACGCTTAC---AAGGCCATCAAGTACGAGGAGTACGCCAGCTCAGCTTCACCGCT 1342
DB 194 CGAGCTTCTACTGATCGGCGCGGCGACGCTGGATCGCGCCAAAGCGCGGCGGTC 253
QY 1343 CCGTCCGCGAAGGAGGCGCCACCCACACCTTCGACGTCAGCGCGCGCCACCTTCGTA 1402
DB 254 TGCTGACGACCTGGCGGTGGCGGACTACAGCGGCTACATCTGCGGCTGCTCGCGGAGG 313
QY 1403 CCGCCA 1408
DB 314 CCGCCA 319

RESULT 12
CNS01VTG      773 bp DNA linear GSS 12-MAY-2000
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence 17 end of clone
199C24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL169349.1 GI:7807506
VERSION
GSS: genome survey sequence.
KEYWORDS
Tetraodon nigroviridis.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 773)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W., and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL
2 (bases 1 to 773)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
JOURNAL
3 (bases 1 to 773)
Genoscope.
REFERENCE
Direct Submission
AUTHORS
Submitted (12-APR-2000)
JOURNAL
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
COMMENT
Location/Qualifiers
FEATURES
source
1..773
/organism="Tetraodon nigroviridis"

```

```

/db_xref="taxon:99883"
/clone="199C24"
/clone_lib="G"
/notes="Genoscope sequence ID : C0AG199B12LPI-end : T7"
BASE COUNT      210 a 265 c 89 g 184 t 45 others
ORIGIN
Query Match      3.8%; Score 75.4; DB 17; Length 773;
Best Local Similarity 43.1%; Pred. No. 1.5e-05;
Matches 259; Conservative 18; Mismatches 324; Indels 0; Gaps 0;

QY 1374 CTTGCGAGCTCAGCGCGCTACCTTCGTGACCGCCACCTCTACTGGACACGCGGTCCAG 1433
DB 166 CTGCTACTACTACWACTGCTACTACTAGTGTGWACHWGTCTACTGCTACTGCTACTA 225
QY 1434 CGACATCGACCTCTACTCTTACGACCCGACGAGGAGGAGGTTCCTACTACACCGC 1493
DB 226 CTAGTCTACTACTGCTACTGCTACTACTACTACWAGTGTCTACTGCTACTGCTACTG 285
QY 1494 CTACTAGGCTTCGGAGAGGCTGCGTACTACACCGCCGACCGGAACTGACGCGGTCAA 1553
DB 286 CTACWACTGCTACTACTGCTACTGCTACTACTACTACTACTACTACTACTACTACTA 345
QY 1554 GGTGCTCAGCTACAGGCGCTGCGGACTACGAGGTGCGTGCACGACGCGGAGCT 1613
DB 346 CTAGTCTACTACTACTGCTACTGCTACTACTACTACTACTACTACTACTACTACTACTA 405
QY 1614 CAGCCAGTCCGGCGCGCCAAACCGAATCCAAACCCGAAACCCGACCGACAC 1673
DB 406 CTGCACTACWCGCACTGCTCTCTCTACAGWACTACTACCAACWACCACTACTACRHA 465
QY 1674 CGACACCGAGCTTCACCGCTTCCGTTAAGGACTACTGGGACACCGACGACCTTAC 1733
DB 466 CWRCACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 525
QY 1734 CATGACGCTCAGAGGCTGCGCCACAGATACCGGTGACCTGACCTGCTACTTCTTA 1793
DB 526 CTACTACTACTAGWGTCTACTACTACTACTACTACTACTACTACTACTACTACTACT 585
QY 1794 CAAGGACCTCGACCTCTACTCTCTACGACCCCAACGCGCACTCTGTCAGAGTCCAGGTC 1853
DB 586 CTACWACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 645
QY 1854 GAGCAACAGCTACGACGAGCTGCTAGTACGCAACCGCCGCGGAACTGCGGTCTCT 1913
DB 646 AACATATCTGCTACTCTGCTACTACTACTACTACTACTACTACTACTACTACTACTA 705
QY 1914 GGTCTAGCGCTACAGCACTACTGCTGCGGCGGACTACAGCTCAAGCGCGCTGCTACTA 1973
DB 706 CTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 765
QY 1974 C 1974
DB 766 C 766

```

```

RESULT 13
BG608096      627 bp mRNA linear EST 17-APR-2001
LOCUS
DEFINITION
WHE2496_H02_004AF1 Triticum monococcum early reproductive apex cDNA
Library Triticum monococcum cDNA clone WHE2496_H02_004, mRNA
sequence.
ACCESSION
BG608096
VERSION
BG608096.1 GI:1558079
KEYWORDS
EST.
SOURCE
Triticum monococcum
ORGANISM
Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
Triticeae; Trit. Gen.
1 (bases 1 to 627)
REFERENCE
Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P.S., Heia
, C.C., Kang, Y., Luo, G.R., Miller, R., Rausch, C.J., Sealon, C.L.,

```

TITLE Stamova, B. and Tong, J.C.
The structure and function of the expressed portion of the wheat
monococcum - Early reproductive apex cDNA library from Triticum
monococcum
Unpublished (2001)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5103595773
Fax: 5105595818
Email: oanderson@wp.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Strategene SK primer.
Location/Qualifiers
1. 627
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone_lib="WHE2496_H02_004"
/tissue_type="Early reproductive apex"
/dev_stage="Seven week-old plants"
/lab_host="E. coli XL0LR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; The tissue, total RNA; and
poly(A) RNA were prepared from apex at double-ridge stage
to terminal spikelet stage during transition from
vegetative state to flower state, a cDNA library was made,
and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova,
J. Dubcovsky). Plasmid DNA preparations and cDNA
sequencing were performed in the OD Anderson lab (a-1
other authors)."

FEATURES
Source

BASE COUNT 142 a 247 c 173 g 65 t
ORIGIN
Query Match 3.88; Score 75; DB 12; Length 627;
Best Local Similarity 49.38; Pred. No. 1.7e-05;
Matches 258; Conservative 0; Mismatches 255; Indels 10; Gaps 2;
QY 1367 CCACACCTTCACGTCAGCGGCCACCTTCGTGACCGCCACCTCTACTGGGACAGG 1426
DB 76 CCAGCTTCACGCTCTCCCTCGGCGACTCCTGCTCCCGCGGACGCCGCCCGCC 135
QY 1427 GCTGAGGACATCGACCTCTACCTCTAGACCCCGACGCGGACGAGTTGACTACT 1486
DB 136 GCACCGGACCGCGCGCGGCTCTCTCCCGCTCCGAGCGGCCACCAACGCGCG 195
QY 1487 ACACCGCTTACTGCGAGAGCTGCGCTACTACACCGGACCGCGGACCTGGA 1546
DB 196 GAACCGGATTCGCGCGCGCGGACGAGTACCGGTACCGCTCAACGCGCGCGCTC 255
QY 1547 CGTCAAGTGTGCTAGTACAGGCGCGGCACTACGAGTTCAGCTGCTGAGGAGC 1605
DB 256 CGGCAAGCTCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
QY 1607 GAGGCTCAGCAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1666
DB 313 TCAGCTTCG 372
QY 1667 CGACCGGACCGACGACCTTCACCGGTTCCGTTTACGACTTACTGGGACCGACG 1726
DB 373 CCAAGCGGACGACGACCTTACCGGTTCCGTTTACGACTTACTGGGACCGACG 1736
QY 1727 CTTTACGATGACGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779
DB 433 GGGGCTACGACGACCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1780
QY 1780 TTCGATCTTCTTACGACGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1839

Db 493 ATCGGACG 552
QY 1840 GACAGTTCAGCTCGGACGACGACGACGACGACGACGACGACGACGACGACG 1882
Db 553 AAGAAGGCG 595
RESULT 14
AW453236
LOCUS
DEFINITION 60033P06.y1.660. Mixed stages of anther and pollen Zea mays cDNA,
mRNA sequence. 559 bp mRNA linear EST 17-FEB-2000
ACCESSION AW453236
VERSION AW453236.1 GI:6994022
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Vascularophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 559)
AUTHORS Walbot V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V.
Department of Biological Sciences
Stanford University
855 California Ave.
Palo Alto, CA 94304, USA
Tel: 650 723 2223
Fax: 650 725 8323
Email: walbot@stanford.edu
Plate: 660033
Location/Qualifiers
1. 559
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="Whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XLOLR"
/note="Organ: anthers; Vector: Lambda Zap; Site 1: EcoRI;
Site 2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 106 a 198 c 177 g 78 t
ORIGIN

Query Match 3.74; Score 73.6; DB 10; Length 559;
Best Local Similarity 50.04; Pred. No. 3.3e-05;
Matches 184; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
QY 1017 CGAAGGCTCAAGCGGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076
DB 192 CGGAGTGTGACG 251
QY 1077 CGGAACGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1136
DB 252 GAGGTCG 311
QY 1137 GCGCAGCG 1196
DB 312 GCGCAGCG 371
QY 1197 CCGCAGCG 1356
DB 372 CAGCG 431
QY 1257 AGCGGACG 1316
DB 432 CCAGATCATGACGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 491

Wed Nov 6 14:29:39 2002

QY 1317 CTAGCCGAAGTCACTTACCGGCTCGTCCGCGGAGGAGGCGCCACACACTT 1376
DB 492 TGTCCCGGCGAAGCGTACAGCCCGGCTGTGTACGACCTCGCGCTCCACGACTAGC 551
QY 1377 CGAGGTCA 1384.
DB 552 CGGCTACA 559

RESULT 15
B0660503
LOCUS
DEFINITION
B0660503
VERSION
B0660503.1 GI:21801636
KEYWORDS
EST.
SOURCE
Hordeum vulgare.
ORGANISM
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 607)
Michael, W., Weschke, W., Pleissner, K.-P. and Graner, A.
EST sequencing and analysis in barley
Unpublished (2000)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 607 Std Error: 0.00
Plate: 2 Row: 0 Column: 5
Seq primer: M13uni.
Location/Qualifiers
1. .607
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="GABI:136895"
/db_xref="taxon:4513"
/clone="H102005"
/clone_lib="H1"
/tissue="female inflorescences"
/dev_stage="female inflorescences (approx. 3 mm in size)"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

BASE COUNT 132 a 251 c 162 g 62 t
ORIGIN

Query Match 3.7%; Score 73.6; DB 14; Length 607;
Best Local Similarity 48.4%; Pred No. 3.4e-05;
Matches 236; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 1397 TCGTGCACCGCCACCTCTACTGGGACCGGCTCGAGCGGACGACCTTACCTCTACG 1456
DB 120 TCCAGAGCGCCACGCTTCTCGCGGCTCTCCCTCGGCGACTCCATCTGGTCTCCCGCGG 179
QY 1457 ACCCAACGGAACGAGGTTGACTTCTCTACACCGGCTACTAGCGCTTCGAGAGTGG 1516
DB 180 ACAACAGCGCCCGCGCGCGCAAGCGGACCGCGGCTCTTCTCCCGCTCCCGCGCG 239
QY 1517 GCTACTACACCGCGCGGACCTGGACGCTCAAGGTGCTGCTACAGCTACAAGGCGCG 1576
DB 240 CGGCCCAACAC---AACGCCCTTCCCAAGCCCAACACAGCTCAACAGCGCGCGGCTCA 296

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

ON nucleic - nucleic search, using sw model

Run on: November 2, 2002, 00:34:13 ; Search time 281 Seconds
(without alignments)
15844.131 Million cell updates/sec

Title: US-09-841-553-2

Perfect score: 1777
Sequence: 1 ATGAAGAGGTAGTGCTGT.....AGCGCGTCTACTACGGG 1977

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being prioritized and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1377	100.0	1977	18	AAT85667
2	1377	100.0	1977	20	AA05926
3	1360.8	68.8	1977	18	AAT85669
4	1155	58.9	1859	18	AAT85677
5	833.2	42.1	898	17	AAT08131
6	806.8	40.8	1464	18	AAT85679
7	750.2	37.9	1962	18	AAT85695
8	750.2	37.9	1962	20	AA05929
9	692.8	35.0	1566	18	AAT85668

10	550.4	28.3	1236	20	AA05920
11	532.8	26.9	564	12	AAT08134
12	532.8	26.9	564	18	AAT85676
13	531.6	8.7	2539	18	AAT61454
14	519.6	8.7	2809	18	AAT61455
15	518.6	8.7	2121	23	ABL54900
16	510.6	7.6	546	23	ABL5453
17	505.4	7.4	507	18	AAV03761
18	425.4	6.3	1140	19	AAV41733
19	424	6.3	145	17	AAT08133
20	424	6.3	145	19	AAT85675
21	415.4	5.8	985	24	ABQ44815
22	415.4	5.8	1311	24	ABQ44817
23	415.2	5.8	319630	24	ABQ67800
24	413.2	5.7	1296	24	ABQ67194
25	413.2	5.7	1296	24	ABQ67797
26	413.2	5.7	4403765	22	AAI99683
27	412	5.7	65140	22	AAI71184
28	412	5.7	125401	22	AAI71186
29	411.8	5.7	1635	22	ABA49946
30	411.8	5.7	1635	22	ABA67865
31	411.8	5.7	1635	22	ABA34921
32	411.8	5.7	1635	22	AAI62270
33	411.8	5.7	1635	22	AAI42018
34	411.8	5.7	1635	22	AAI22780
35	411.8	5.7	1635	22	AAI18082
36	411.8	5.7	1635	22	AAI08454
37	411.8	5.7	1635	22	ABA44807
38	411.8	5.7	1635	22	ABA44807
39	411.8	5.7	1635	22	ABA55261
40	411.8	5.7	1635	22	ABA25005
41	411.8	5.7	1635	22	AAI03814
42	411.8	5.7	1635	22	AAI03814
43	411.8	5.7	1635	22	AAI28370
44	411.8	5.7	1635	22	AAI13356
45	411.8	5.7	1635	22	AAI34918
					AAI03446

ALIGNMENTS

RESULT 1
AAT85667
ID AAT85667 standard; DNA; 1977 BP.
XX
AC AAT85667;
XX
DT 20-APR-1998 (first entry)
XX
DE Thermococcus protease coding sequence.
XX
KW Protease; research reagent; thermal stability: thermococcus celer; aa.
XX
OS Thermococcus celer DSM-2
XX
PN WO9721823-A1.
PD 19-JUN-1997
XX

PF 07-NOV-1996; 96WO-JF0324
XX
PR 12-DEC-1995; 95JP-032324F
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Asada K, Kato I, Mita M, Morishita M, Takakura H;
XX Tsunawasa S, Yamamoto K;
XX HPI; 1997-332794/30.
XX P-PSDB; AAW24121.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
Pyrococcus strains - have extremely high thermal stability and are

Hyperthermostable
DNA sequence. Pyr
Thermococcus prote
Streptomyces virid
DipA-mel chimeric
T. yonsei subcl
Pyrrococcus proteas
Codon-optimized RA
DNA sequence. Pyr
Thermococcus prote
Oligonucleotide fo
Oligonucleotide fo
Listeria innocua D
Listeria innocua D
Mycobacterium tube
Streptomyces nous
Human breast cell
Human foetal liver
Probe #13387 for g
Human brain expres
Human bone marrow
Probe #12713 for g
Probe #16768 used t
Probe #8445 used t
Human genome-deriv
Human breast cell
Probe #3471 for ge
Human brain expres
Human bone marrow
Probe #3489 for ge
Probe #3604 used t
Probe #3437 used t

Db 1921 GCCTACAGCACCCTACGGCTGGCGGGACTACCAAGCTCAAGGCCGTCGTCTACTACCGG 1977

RESULT 2

AAX05926

ID AAX05926 standard; DNA: 1977 BP.

XX
AC AAX05926;

XX
DT 05-MAY-1999

00-MAY-1999 (First)

DE WO9856926 Seq ID 11.
XX

KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin; additional data

addictive; drug; was

~~DS Thermococcus~~

PN
N
X

PD 17-DEC-1998.

04-JUN-1998; 98WO-JP02465

10-JUN-1997 03:00:00

10-JUN-1997; 9/JP-0151969.

(TAKI) TAKARA SHUZO

I Asada K, Kato I,
X

WPI; 1999-080907/07.

P-PSDB; AAW94840.

Recombinant hyperthermostable protease from *Pyrococcus furiosus* and gene encoding it for

and gene encoding it, for large scale production of the protease for industrial use.

Disclosure; page 53-54; 82pp; Japanese

The invention relates to a Japanese.

thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has an inversion related to a hyperthermostable protease derived from a

working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 (optimum 6-8), and retains more than 90% of its activity.

at 95 deg C. The invention also provides gene sequences encoding a polypeptide of formula (I) or (II).

polypeptide of formula $\text{Sig-Ala-Gly-Gly-Asn-Pro}$, where Sig is a signal peptide from subtilisin, and Pro is the above protease. Host cells

(especially *Bacillus* strains) transformed with vectors comprising the genes are used for the recombinant production of the enzyme.

hypothermostable protease which can be prepared in quantity suitable for industrial use can be used as a substitute for the protease.

for industrial use, can be used as an additive for drugs, washing agents and foodstuffs and for chemical synthesis.

Sequence 1977 BP: 453 A: 659 C: 554 G: 311 T: 0 other.

Query Match

QY	241	CTCATGGGGCCCGAGCTCTAGTACTCTCTACAAGATAATCTCTGCTGCGCGGTAAATA	300
DB	241	CTCATGGGGCCCGAGGTCAGTCTCTACAAGATAATCTCTGCTGCGCGGTAAATA	300
QY	301	AAGCCAGGAGCCTTCTCTGATCTCGGGGCATATAGACACGGTTACTCTGGGTACACA	360
DB	301	AAGCCAGGAGCCTTCTCTGATCTCGGGGCATATAGACACGGTTACTCTGGGTACACA	360
QY	361	AGGCTCTCGGCATTAAGTTCTATACAGAGGATTTACAGGTTTCAAGTTGACAGCGCACT	420
DB	361	AGGCTCTCGGCATTAAGTTCTATACAGAGGATTTACAGGTTTCAAGTTGACAGCGCACT	420
QY	421	TCCGTCCTCCAGATAGGAGCGGATACCGTCTTGAATCTCCCTGGCTACAGCGGAAGCGGT	480
DB	421	TCCGTCCTCCAGATAGGAGCGGATACCGTCTTGAATCTCCCTGGCTACAGCGGAAGCGGT	480
QY	481	GTGGTGGTGCCATCTGCTCATACAGGATATAGACGCCCAACCCCGCATCTGAAGGGCAAG	540
DB	481	GTGGTGGTGCCATCTGCTCATACAGGATATAGACGCCCAACCCCGCATCTGAAGGGCAAG	540
QY	541	GTCAATAGGCTGTACGACACCGCTCAAGCGCAGTGCACCGGCTACGATGACCAAGGACAC	600
DB	541	GTCAATAGGCTGTACGACACCGCTCAAGCGCAGTGCACCGGCTACGATGACCAAGGACAC	600
QY	601	GGAAAGCCAGCTTGGGGTATCTGTCGGCAACCGGCAAGCTTAACTCCAGCTACATAGGC	660
DB	601	GGAAAGCCAGCTTGGGGTATCTGTCGGCAACCGGCAAGCTTAACTCCAGCTACATAGGC	660
QY	661	GTGCCCCCGCGCGAAGCTCTGTGGGCTCAAGGTTCTCGGTGCGGACGGTTTCGGGAAGC	720
DB	661	GTGCCCCCGCGCGAAGCTCTGTGGGCTCAAGGTTCTCGGTGCGGACGGTTTCGGGAAGC	720
QY	721	GTCTCCACCATCATCGGCGTGTGACTGGGTGCTCGAAGCAAGACAAAGTACGGGATA	780
DB	721	GTCTCCACCATCATCGGCGTGTGACTGGGTGCTCGAAGCAAGACAAAGTACGGGATA	780
QY	781	AGGCTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGAACCGAGCTCCCTCAGT	840
DB	781	AGGCTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGAACCGAGCTCCCTCAGT	840
QY	841	CAGGCGCTCAACAGCCCTCTGCAGCCCGGTTATAGTAGTCTGCGTCCGCGCGGCAACAGC	900
DB	841	CAGGCGCTCAACAGCCCTCTGCAGCCCGGTTATAGTAGTCTGCGTCCGCGCGGCAACAGC	900
QY	901	GGGCGGAACACCTTACACGCTACCGCTACCGCCCGCGCGGAGTCAATACCGTCGCT	960
DB	901	GGGCGGAACACCTTACACGCTACCGCTACCGCCCGCGCGGAGTCAATACCGTCGCT	960
QY	961	GCATCTCAGACGAGCAACATCGGCGAGCTTCTCCAGCAGGGAGCCGACCGCGACGGA	1020
DB	961	GCATCTCAGACGAGCAACATCGGCGAGCTTCTCCAGCAGGGAGCCGACCGCGACGGA	1020
QY	1021	AGGCTCAAGCCGGAAGTCTGCGCCCGCGGTTGACATCATAGCCCGCGCGGCAACGGGA	1080
DB	1021	AGGCTCAAGCCGGAAGTCTGCGCCCGCGGTTGACATCATAGCCCGCGCGGCAACGGGA	1080
QY	1081	ACCAGCATGGGCACCCGCTAAGACTACTACACCAAGGCTCTTGGAAACGATGGCC	1140
DB	1081	ACCAGCATGGGCACCCGCTAAGACTACTACACCAAGGCTCTTGGAAACGATGGCC	1140
QY	1141	ACCCGCGAGTTTGGGCTTGGCGGCTGCTATCTCCAGGCGCCACCGGAGTGGACCCCG	1200
DB	1141	ACCCGCGAGTTTGGGCTTGGCGGCTGCTATCTCCAGGCGCCACCGGAGTGGACCCCG	1200
QY	1201	GACAGGTTGAAGACCGGCTCATCTGAGACCCGCGACATAGTCCGCCCGCAGGAGTACGC	1260
DB	1201	GACAGGTTGAAGACCGGCTCATCTGAGACCCGCGACATAGTCCGCCCGCAGGAGTACGC	1260
QY	1261	GACATCGGCTACGTTGGGCTAAGGTTGAGGTTCTACAGGCCATAGTACGACGACTAC	1320
DB	1261	GACATCGGCTACGTTGGGCTAAGGTTGAGGTTCTACAGGCCATAGTACGACGACTAC	1320

QY 1321 GCCAAGCTCACCTTACCGGCTCGCTCGGACAAAGGAAGCGCCACCCACACCTTCGAC 1380
DB 1321 GCCAAGCTCACCTTACCGGCTCGCTCGGACAAAGGAAGCGCCACCCACACCTTCGAC 1380
QY 1381 GTCAGCGCGCGACCTTCTGTCAGCGCCACCTCTACTGGGACAGCGGCTCGAGGACATC 1440
DB 1381 GTCAGCGCGCGACCTTCTGTCAGCGCCACCTCTACTGGGACAGCGGCTCGAGGACATC 1440
QY 1441 GACCTCTACCTTACGACCCCAAGGAGAGAGAGGTTGACTTACTCTACACCGCTTACTAC 1500
DB 1441 GACCTCTACCTTACGACCCCAAGGAGAGAGGTTGACTTACTCTACACCGCTTACTAC 1500
QY 1501 GCTTTCAGAGGCTCGGCTTACTACAAACCGACCGCGGAACTGAGCGGTCAAGGTCGTC 1560
DB 1501 GCTTTCAGAGGCTCGGCTTACTACAAACCGACCGCGGAACTGAGCGGTCAAGGTCGTC 1560
QY 1561 AGCTACAAAGGCGCGGCGGAACTACCGGTGAGCTGCTGAGGAGCGGAGCTCAGCTAC 1620
DB 1561 AGCTACAAAGGCGCGGCGGAACTACCGGTGAGCTGCTGAGGAGCGGAGCTCAGCTAC 1620
QY 1621 TCCGCGCGGCGCAACCGGAACTCCAAACCGGAACTGAGCGGAGCGGAGCTCAGCTAC 1680
DB 1621 TCCGCGCGGCGCAACCGGAACTCCAAACCGGAACTGAGCGGAGCGGAGCTCAGCTAC 1680
QY 1681 CAGACCTTACCGGCTTCCGTTTACGACTTACTGGGACAGCGGAGCGGAGCTCAGCTAC 1740
DB 1681 CAGACCTTACCGGCTTCCGTTTACGACTTACTGGGACAGCGGAGCGGAGCTCAGCTAC 1740
QY 1741 GTCACAGCGGTGCCAGGAGTACCGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1800
DB 1741 GTCACAGCGGTGCCAGGAGTACCGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1800
QY 1801 CTCGACCTTACCTTACGACCGCAACCGGAACTCGTTGACAGGTGACCGGAGCGGAG 1860
DB 1801 CTCGACCTTACCTTACGACCGCAACCGGAACTCGTTGACAGGTGACCGGAGCGGAG 1860
QY 1861 AGCTAGGAGCGGTGAGTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1920
DB 1861 AGCTAGGAGCGGTGAGTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1920
QY 1921 GCTTACAGACCTTACGCTGGGCGGAGTACAGCTTACAGCTTACAGCTTACAGCTTAC 1980
DB 1921 GCTTACAGACCTTACGCTGGGCGGAGTACAGCTTACAGCTTACAGCTTACAGCTTAC 1980

RESULT 3
KAT85669
ID KAT85669 standard; DNA, 1977 BP.
AC KAT85669;
XX KAT85669;
DT 20-APR-1998 (first entry)
XX Protease coding sequence.
XX Protease, research reagent; thermal stability: ss.
XX Synthetic.
XX W09721823-AL.
XX PD 19-JUN-1997
XX 07-NOV-1996; 96W0-JP03253.
XX 12-DEC-1995; 95JP-0323285.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Asada K, Kato I, Mitra M, Morishita M, Takakura H;
XX Tsunashima S, Yamamoto K;
XX WPI; 1997-332794/30.

DR P-PSDB; AAN24123.
XX Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are
PT useful industrially and as research reagents
XX Claim 11; Page 95-97; 159pp; Japanese.
XX This sequence represents the coding sequence for a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries.
XX Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 other;
SQ Query Match 68.6%; Score 1360.8; DB 18; Length 1977;
Best Local Similarity 83.9%; Pred. No. 7.3e-242;
Matches 1573; Conservative 0; Mismatches 292; Indels 9; Gaps 3;
QY 110 AGAAGACTACGAGCTGCTACCGCGGAGCTGTTCAAGAAAGTCCAGAGGATGACTGGA 169
DB 107 AGAAGACTATGCTGCTACCGCGGAGCTGTTCAAGAAAGTCCAGAGGATGACTGGA 166
QY 170 ACCAGAGTGGACACCGTCAATGTTCCGGAGCTACGAGACAGGACAGGCGGCTTA 229
DB 167 AGAGAAATCAGCAGTAATGTTGTTGAAGAACATAGGGAAGAAATTCAGTAA 226
QY 230 AGTACTGAGGCTGATGGCGGCGGAGCTGCTGCTGATGCGGCGGATGATA---GACACGGGTT 346
DB 227 GAGTCTTGTGTTAAGGTC---AAGATTAGGTATGTGACCATATATATCCCGCAATAG 286
QY 290 CGGTTAAATAAAGCGGAGCGGAGCTTCTGCTGATGCGGCGGATGATA---GACACGGGTT 346
DB 287 CTGCGGACTTAAGGTTAGAGATTTACTAGTCACTCAGGTTTAAACAGGGGTTAAAGCTA 346
QY 347 ACTTGGTAAACAAAGGCTTGGGCTAAGTTCATACAGGAGGTTTACAGGTTTACAGG 406
DB 347 AGTCTTACAGGTTGAGGTTTATCAGGAGAGCTACAAAGTTACGTTTTCAGCAGAAATAG 406
QY 407 TTGAGCGGCGGAGCTTCCGCTGCTGAGTAGGAGGCGGATACCGTCTGAGAACTCCCTCGGCT 466
DB 407 AAGGACTGATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
QY 467 ACAGCGAAGGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 526
DB 464 ATGATGTTCTGGGATCAGCAAGGAAATTAATGACACTGGAATGACGCTTCTATCCAG 523
QY 527 ATCTGAAGGCGAGGCTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 586
DB 524 ATCTCAGGAGAAAGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 583
QY 587 ATGACCGGAGCAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
DB 584 ATGACCATGACATGGAACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
QY 644 ACTCCGAGTACATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
DB 644 ATGCGAGTACAGGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
QY 704 CCGAGGCTTCCGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763
DB 704 CCGATGCTTCCGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763
QY 764 AGGACAGTACGAGGATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
DB 764 AAGATAAGTACGAAATTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
QY 824 GAACCGGCTTCCGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
DB 824 GAACCGGCTTCCGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
QY 884 TCGCGCGGCGGAGCAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943

Wed Nov 6 14:29:34 2002

QY	1173	CCTCAGGCGCCCGAGCTGGACCCGCGACAGGTGAAGACCGCCCTCATCGAGCCG	1232	FT	misc_difference 587	a	"base n at position 587 is not identified in the specification"
DB	361	CTCAGGCGCCCGAGCTGGACCCGCGACAGGTGAAGACCGCCCTCATCGAGCCG	420	FT	misc_difference 718	b	"base n at position 718 is not identified in the specification"
QY	1233	CGACATAGTCCGCGCCCGAGGAGATAGCGGACATCGCCTACGCTGCGGTAGGTGAACCT	1292	FT	misc_difference 784	c	"base n at position 784 is not identified in the specification"
DB	421	CGACATAGTCCGCGCCCGAGGAGATAGCGGACATCGCCTACGCTGCGGTAGGTGAACCT	480	FT	misc_difference 796	d	"base n at position 796 is not identified in the specification"
QY	1293	CTACAAGGCCATCAAGTACGAGCATACGCCAAGCTCACTTACCGGCTCGTCCGCGA	1352	FT	misc_difference 801	e	"base n at position 801 is not identified in the specification"
DB	481	CTACAAGGCCATCAAGTACGAGCATACGCCAAGCTCACTTACCGGCTCGTCCGCGA	540	FT	misc_difference 815	f	"base n at position 815 is not identified in the specification"
QY	1353	CAAGGAGGCGCCACACACCTTCGAGCTGACGGCGCCACCTTCGTCGAGCCGACCT	1412	FT	misc_difference 859	g	"base n at position 859 is not identified in the specification"
DB	541	CAAGGAGGCGCCACACACCTTCGAGCTGACGGCGCCACCTTCGTCGAGCCGACCT	600	FT	misc_difference 878	h	"base n at position 878 is not identified in the specification"
QY	1413	CTACTGGGACACCGGCTCGAGGACATCGACCTTACCTCTACGACCCCAACCGGAAGA	1472	FT	misc_difference 887	i	"base n at position 887 is not identified in the specification"
DB	601	CTACTGGGACACCGGCTCGAGGACATCGACCTTACCTCTACGACCCCAACCGGAAGA	660	FT	misc_difference 890	j	"base n at position 890 is not identified in the specification"
QY	1473	GTTGACTACTCTCTACACCGCTACTACGGCTTCGAGAGGTTCGGCTACTACACCCGAC	1532	FT	misc_difference 896	k	"base n at position 896 is not identified in the specification"
DB	661	GTTGACTACTCTCTACACCGCTACTACGGCTTCGAGAGGTTCGGCTACTACACCCGAC	720	FT			
QY	1533	CGCGGACCTGAGCTCAAGTCTGAGCTCAAGGCGCGGCGGAGTCACTACGAGTCA	1592	FT			
DB	721	CGCGGACCTGAGCTCAAGTCTGAGCTCAAGGCGCGGCGGAGTCACTACGAGTCA	780	FT			
QY	1593	CGTCTGAGGACCGGAGCTTACGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1652	FT			
DB	781	CGTCTGAGGACCGGAGCTTACGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	840	FT			
QY	1653	CCGGAACCCAGCCCGGACCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1712	FT			
DB	841	CCGGAACCCAGCCCGGACCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	900	FT			
QY	1713	GGACACGAGGACCTTACCAATGAAGTCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1772	FT			
DB	901	GGACACGAGGACCTTACCAATGAAGTCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	960	FT			
QY	1773	CGTCTGAGGACCGGAGCTTACGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1832	FT			
DB	961	CGTCTGAGGACCGGAGCTTACGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG	1020	FT			
QY	1833	CCTCGTTGACAGGTCACGTCGAGCAACAGCTACGAGACGTCGAGTACGACCCGCG	1892	FT			
DB	1021	CCTCGTTGACAGGTCACGTCGAGCAACAGCTACGAGACGTCGAGTACGACCCGCG	1080	FT			
QY	1893	CCGCGGACCTGAGCTTCTCGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTAC	1952	FT			
DB	1081	CCGCGGACCTGAGCTTCTCGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTAC	1140	FT			
QY	1953	GCTCAAGCGCGTGGCTTACTACGGG	1977	FT			
DB	1141	GCTCAAGCGCGTGGCTTACTACGGG	1165	FT			
RESULT 5				PN	WO9534645-A1.		
ART08131				XX	21-DEC-1995.		
ID	ART08131	standard:	DNA: 898 bp.	PD	05-JUN-1995; 95WO-JP01095		
AC	XX			PR	26-JUL-1994; 94JP-0173517		
XX	XX			PR	13-JUN-1994; 94JP-013046		
AC	ART08131;			XX	(TAKI) TAKARA SHUZO CO LTD.		
XX	10-MAY-1996 (first entry)			XX	Asada K, Kato I, Mitsu N, Morishita M, Tanasawa S;		
XX	Hyperthermostable protease gene.			XX	Yamamoto K;		
XX	Protease; hyperthermostable; thermostability; ss.			DR	WPI: 1996-049674/05.		
XX	Pyrococcus furiosus.			DR	P-PSDB: AAR87009.		
XX	Key	Location/Qualifiers		XX	Pyrococcus furiosus hyperthermostable protease gene - useful for recombinant prodn. of hyperthermostable protease		
XX				XX	Example 3: Page 52-53; 85p; Japanese.		
XX				XX	The invention relates to the hyperthermostable protease of Pyrococcus furiosus and its prodn. as a recombinant protein		
XX				XX	in transformants using a vector carrying the protease gene (AAT08141). A genomic DNA sequence of the invention is given in		
XX				XX	AAT08131 and its encoded protein in AAR87009.		

512 CTTCTCATCCAGATCTCCAAAGAAAGTAATTTGGGTGGGTAGATTTTGTCAATGGTAGCA 591

575 CGACCCCTCAGATGACAGGACGACGAGAACCCAGCTTGGGGTATTCGTTCCCGGAACCG 634

572 GTTATCATACGATGACCATGACATGGAATCATGTAGCTTCAATAGCAGCTGTGACTG 631

635 GCAGCG--TTAACTCCCGCTACATAGGGGTGCGCCCGGCGGAAGCTCGTGGCGGTCA- 691

632 GAGCAGCAAGTAATGCAAGTACAAAGGAAATGGCTCAGAGCTAAGCTGGCGGGAATTA 691

692 AGGTTCTCGGTGCGGACGCTTCGGGAAGCGCTCCACCATCATCGCGGTGTTCAGCTGG 751

692 AGGTTCTAGGTGCGGATGGTTCTCGGAAGCATATCTACTATAATTAAGGGAGTGTAGTGG 751

752 TCGTCCCGAACAAGGACAAAGTACGGGATTAAGGGTATCAAGCTCTCTCCGCGCTCTCCC 811

752 CGTTGATACAAAGATTAAGTACGGAATTAAGGTCATAATCTTCTCTGTGTTCAAGCC 811

812 AGAGTCCGCGAGAAACCGACTCCCTCAGTCAGCGCGTCAACAAAGCGCTGGAGACCGCGTA 871

812 AGAGTCTAGATGGTACTGACGCTCTAAGTCAGGCTTTAAGTCAGGCTTTAAGTCAGCTGAT 871

872 TAGTAGTCTGGGTGCGCGCGGCAACAGCGGGCGAACAACCTACACGCTGGCTCACCCTG 931

872 TAGTTGTTGTGGTTTCCCGTGTGAAACAGTGGAGCTTACCAAGTATACANTCGTTCTCAG 931

932 CGCGCGAGCAAGGTCATAAAGCTGCGGTGAGTTGACAGCAAGCAACATCGCAGCT 991

932 CAGCTCCAGCAAGATTTATTACAGTTGGAGCGCTTGACAAGTATGATGTTATTACAACT 991

992 TCTCAGCAGGGGACCGACCGGGGACGAAGGCTCAAGCGCAAGCTGTCGCGCCCGCTCG 1051

992 TCTCAAGCAGAGGGGCAACTGCAAGCGGAGGCTTAAGCCTTAAGCTTGTTGCTCCAGGAA 1051

1052 TTGACATCATAGCCCGCGCGCGAGGAGCAACAGCTATGGGCAACCGCATTAACGACTACT 1111

1052 ACTGGATAATTCGTCGCCAGAGCAAGTGGAACTAGCATGAGGTCAAGCTTAATGACTATT 1111

1112 ACACAAGGCTCTTGGAAACAGCATGGCCACCCCGCACGTTTCGGGCGCTGGCGCGCTCT 1172

1112 ACACAGCAGCTCTGGGACATCAATGGCAACTCTCTACGTFAGCTGGTATTCGACGCTCT 1172

1172 TCTTCAGGCCACCGGAGCTGGACCCGGAAGGTGAAGACCGCCCTCATCGAGACCG 1231

1172 TGCTCCAGGACACCGGAGCTGAGCTGACAGACAGTAAAGACAGCCCTCATAGAAACG 1231

1232 CGGACATAGTCGCGCCCGCAAGAGATACCGGACATCGCTACGGTGGGTAGGTGAACG 1291

1232 CTGATATCTGTAAGACCATGAATAGCCGATATAGCTATAGCTGACGTAGGTAAATG 1291

1292 TCTCAAGGCCATCAAGTACAGCACTACGCCAGCTCACTTCCTACCGGCTCCGTCGCGG 1351

1292 CATCAAGGCTATAAATACGATCACTATATGCAAGCTAGTGTTCATCTGGATATGTCGA 1351

1352 ACAAGGACGCCACCCACACTTCGACGCTCAGCGCGCCACCTCTCTGACCGCCACCC 1411

1352 ACAAGGACGCCAACTCACNGTTCGTTATTAAGCGAGCTTCGTTCTGTAAGTCCACAT 1411

1412 TCTACTGGGACAGGGCTCGAGCCGACATCGACCTCTACCTTAGGACCCCAACCGGAAG 1471

1412 TATATGGGACATATGCAATAGCGACCTGTATCTTTTACTCTAGATCCCATGGGAAC 1471

1472 AGGTTGACTACTCTACCGGCTACTACGGCTTCGAGAGGTCGGGTACTACACCGGA 1531

1472 AGGTTGACTACTCTTACCGGCTACTACTATGATTCGAAAGGTTGGTTATTACAACTGTA 1531

1532 CGCGCGGAACCTGAGCGGTCAAGTCTCAGCTTCAAGGGGCGCGGAACCTACCAAGTCTG 1591

1532 CTATGGGAACATGGCAATTAAGGTTGTAAAGTACAGCGGAAGTCAAACTATCAAGTAG 1591

1592 ACCTGTCAGGACGAGGAGCTTCAACGAGTCCGCGCGGCAACCGGAATCCAAACCCCA 1651

1592 ATGCTAAGTATGATGTTGTTCCCTTTCACAGCTTGAAAG-----TTTCACAT-1636

QY	1652	ACCGAACCCACACCCGACACACGACACCCAGACGCTTACCGGTTCCGTTACGACTACT	1711
Db	1637	CTCCACACCAACCAACAGATTTGCAAGAGACGTTCCAAAGGATCCGATCACTACTACT	1696
QY	1712	GGGACACGAGCGACACCTTCACCATTAACGCTCAACAGCGGTGCCACCAAGATAACCGGTG	1771
Db	1697	ATGACAGGAGCGACACCTTTACAAATGACCGTTAACTCTGGGCTACAAAGATTACTGGAG	1756
QY	1772	ACCTGACCTTCGATCTCTATACAGCACTTCGACCTCTACCTCTACGACCCCAACGGCA	1831
Db	1757	ACCTAGTGTTCGACCAAGCTTCATGATCTTGACCTTTACCTCTACGATCTCAACACAGA	1816
QY	1832	ACCTCGTTGACAGGTCCACGCTGAGCAACACCTACGACAGCTGAGTACGACCAACCCCG	1891
Db	1817	AGCTTGTGATAGATTCGGAGAGTCCCAACAGCTACGACACCTAGATATTAATTAACCCCG	1876
QY	1892	CCCGGGAACCTGGACGCTCTCTGCTAGGCTACAGCACTACGCTGGGCGGACTACC	1951
Db	1877	CCCGAGGACCTGGTACTCTCTAGTATATGCTTACTACACTTACGTTGGGCTTACTACG	1936
QY	1952	AGCTCAAGGCGCTGCTACTACGG	1976
Db	1937	AGCTCAGGCTAACTTATATGG	1961
RESULT 9			
AAT85668			
ID	AAT85668 standard; DNA; 1566 BP.		
XX	AC	AAT85668;	
XX	DT	20-APR-1998 (first entry)	
XX	DE	Pyrococcus furiosus protease coding sequence.	
XX	DE	Protease; research reagent; thermal stability; pyrococcus furiosus; s	
XX	OS	Pyrococcus furiosus DSM-3632.	
XX	Key	Location: /aa:1ifiers	
XX	CDS	1..1566	
FT	FT	/tag=	
FT	FT	/transl_except= (pos: 1282..1284, aa: Xaa)	
FT	FT	/note= "396- Gly, Val"	
XX	PN	W09721823-A1.	
XX	PD	19-JUN-1997.	
XX	PE	07-NOV-1996; 96WO-JP03253	
XX	PR	12-DEC-1995; 95JP-0323285.	
XX	PA	(TAKI) TAKARA SHUZO CO LTD.	
XX	PI	Asada K, Kato I, Mita M, Morishita M, Takakura H;	
XX	PI	Tsunasawa S, Yamamoto K;	
XX	XX	WPI; 1997-332754/30.	
XX	DR	P-PSDB: AAW24122.	
XX	XX	Protease(s) and genes encoding them obtained from Thermococcus and	
XX	XX	Pyrococcus strains - have extremely high thermal stability and are	
XX	XX	useful industrially and as research reagents	
XX	XX	Claim 7; Page 90-91; 155pp; Japanese.	
XX	XX	This sequence represents the coding sequence for the protease from	
XX	XX	Pyrococcus furiosus DSM-3632. This sequence encodes a protease of the	
XX	XX	invention. The protease of the invention have extremely high thermal	
XX	XX	stability. The protease can be used as research reagents, and	
XX	XX	industrially in the food, drug and chemical industries.	

```

XX SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 364 T; 1 other;
Query Match 35.0%; Score 692.8; DB 18; Length 1566;
Best Local Similarity 67.1%; Pred. No. 8.66-119;
Matches 1020; Conservative 0; Mismatches 483; Indels 18; Gaps 2;

Oy 459 CTTCCGCTACGACGAGACCGGTGGTGGTGGTCCATCGTACGAGCGGTATAGACGGA 5:8
Db 60 CTTGGGATATATGGTTCTGGAATCACATAGGAATAATTCACACTGGAAATGACGGTTC 119
Oy 519 CCACCCGATCTGAAGGCGAAGGTCTATAGGCTGAGCGCGGTCAAGCGCAGGTGCGAC 5:8
Db 120 TCATCCAGATCTCCACGAGAAAGTAATGGGTGGTAGATTTTGTCAATGGTAGAGTA 179
Oy 579 CCGCTACCATGACGACGAGCAGACCGACCGTTGCGGGGTATCGTTGCGGGAACCGCAG 638
Db 180 TCCATACCATGACGCTGACATGGAACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGC 239
Oy 639 CG---TTAACTCCAGTACATAGGCGTCCGCCCGCGGGAAGCTCGTGGCGTCAAGGT 695
Db 240 AGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGTAACTGGCGGGAATTAAGT 299
Oy 695 TCTCGGTGCGGCGGTTCGGGAAGCGTCTCCACCATCATCGCGGGGTGTGACTGGGTCT 755
Db 300 TCTAGGTGCGGATGTTTCGGAAGCATATCTACTATATTAAGGAGGTGAGTGGCGGT 359
Oy 755 CCAGAACAGGCAACTACGGGATAGGGTCTCAACCTCTCCCTCGGCTCTCCAGAG 815
Db 360 TCATACAAAGATAGTACGGAATTAAGGTCAATTAATCTTCTTGTTCGAAGCAGAG 419
Oy 815 CTCGAGCGGAACCGACTCCCTCAGTACGAGCGGTCAACACGCGTGGAGCGCGGTATAGT 875
Db 420 CTCAGATGATGACTGACGCTCAAGTCAAGCTGTTAATGACGCGTGGATCTGGATTA 479
Oy 875 AGTCTGGTCCCGCGCGCAACAGCGGCGGAACCTTACACCGTCCGGCTACCGCCGCGC 935
Db 480 TGTGTGTTGGCGTGGAAACACTGACCTTAACAAGTATCAATCGGTTCTCCACAGC 539
Oy 936 CGGAGCAAGCTCATACCGTCCGTGCGTGTGACAGCAACGACATCGCCAGCTCTC 995
Db 540 TGCANGCAAGTTATACAGTTGAGCGGTGACAGGTATGATGTTATACAAGTCTCTC 599
Oy 996 CAGCAGGAGCGACCGCGGACGGAAGCTCAAGCGGGAAGTCTGCGCCCGCGGCTCA 1055
Db 600 AAGCAGAGGCGCACTGACAGCGCGAGGCTTAAGCTCGAGTGTGCTCCAGGAAAGT 659
Oy 1056 CATCATAGCCCGCGCGCCAGCGAGCATGGGACGATGGGACCGCCGATTAAGACTACTAC 1115
Db 660 GATAATGCTGCGAGGCAAGTGGAACTAGCATGGGTCAACCAATTAATGACTATTAC 715
Oy 1116 CAAGCGCTGTGAGACCATGACCCAGCGACGTTTCGGGGTGGCGCGCTCATCC 1175
Db 720 AGCAGCTCTGGGACATCAATGCACTCTCAGTACGTTGGTATTCGAGCCCTTCT 775
Oy 1176 CCAGCGCCACCGAGCTGGACCGCGCAAGAGTGAAGACCGCCCTCATCGAGCGCGCA 1225
Db 780 CCAAGCAGACCGCGAGCTGACAGCAAGTAAAGACAGCCCTCATGAAGACTGCTGA 819
Oy 1236 CATAGTGGCGCCCAAGGAGATACCGGACATCGCTACGGTGGGGTGAAGTGAAGTCTA 1295
Db 840 TATCGTAAAGCCAGTGAATATAGCCGATATAGCCTAGGTGCGAGGTAGGTTAATGATA 899
Oy 1296 CAAGGCCATCAAGTACGACGATACGCAAGCTCACCTTCACCGGCTCGCTCGCGGAC 1355
Db 900 CAAGGCTATAAATACGATATAGCCGATATAGCCTAGGTGCGAGGTAGGTTAATGATA 899
Oy 1355 GGAAGGCCACCGACACTTCGAGCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 1415
Db 960 AGGCAGGCAAACTCAGCACTTGGTATTAAGCGGAGCTTGTGTAAGTGGCAATATA 1019
Oy 1416 CTGGGACACGGGCTCAGCGAGATCGACCTCTACGACCCCAAGCGGAACGAGT 1475

```

```

Db 1020 CTGGACAATGCCAATAGCAACCTTGATCTTACCTCTACGATCCCAATGGAACAGGT 1079
Oy 1476 TGACTACTCTACACCGCTTACTACGGCTTCGAGAAGCTCGGTACTACACCGGACCGC 1575
Db 1080 TGACTACTCTTACACCGCTTACTACTGATTCGAAAGGTGGTTATTACAACCCACTGA 1139
Oy 1536 CGGAACCTGGAGCGTCAAGTTCGTACGTACAAAGGCGGCGGCAAGTACAGGTGACGT 1595
Db 1140 TGAACATGGCAATTAAGCTTGTAGCTACAGCGGAAGTGCAAACTATCAAGTAGATGT 1199
Oy 1596 CGTCAGCGAGCGGAGCGCTCAACAGTCCGGGCGGCGGCAACCGGAATCCAAACCCCAACC 1655
Db 1200 GGTAGTATGTTCCCTTTCACAGCTTGAAG-----TTCAATCTCC 1344
Oy 1656 GAACCCAAACCGGACCGACCGCAACCGCAGACCTTCACCGTTCCTGTTAAGGACTACTGGGA 1715
Db 1245 ACACCGAAGCAACAGTAGAGCGCAAGACGTTCCAAAGNATCGATCTACTACTATGA 1304
Oy 1716 CACGAGCGACACTTCACTTGAAGTCAACAGCGGTGCGCACCAGATACCGGTGACCT 1775
Db 1305 CAGGAGCGACACTTCAATGACCGCTTAACTCTGGGCTCAAAAGATTCTGGAGACT 1364
Oy 1776 GACCTTCGATACTTCTTCAACGACCTCGACCTCTACCTCTACGACCCCAAGCGCAACT 1835
Db 1365 AGTGTTCACCAAGCTTCCATGATGATGATGATGATGATGATGATGATGATGATGAT 1424
Oy 1836 COTTACAGGTCCACCTCGGCAACAGCTACGAGCAGCTCGAGTACGCAACCGCCGCC 1895
Db 1425 TGTAGTAGATCGGAGAGTCCCAACAGCTACGAGACGCTAGATTAATTAATTAATTAAT 1484
Oy 1896 GGGACCTGGACCTTCTTCTGCTACGCTACAGCCTACGCTGCGGCGGAGTACCAGCT 1955
Db 1485 AGGAACCTGCTACTTCTTCTAGTATGCTTCTACTACTACGCTTACGCTTGGGCTTACTACGAGCT 1544
Oy 1956 CAAGCGCGCTCTACTACTAG 1376
Db 1545 GACGCTAAAGTTTATTATGG 1565

```

```

RESULT 10
AX05920
ID AX05920 standard; DNA; 1236 BP.
XX
AC AX05920;
XX
DT 06-MAY-1999 (first entry)
XX
DE Hyperthermostable protease fragment encoding DNA.
XX
KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
additive; drug; washing agent; foodstuff; chemical synthesis; DS.
XX
OS Pyrococcus furiosus.
XX
PN M09856926-A1.
XX
PD 17-DEC-1998.
XX
PF 04-JUN-1998; 98WO-JP02452.
XX
PR 10-JUN-1997; 97JP-015155.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Asada K, Kato I, Morishima M, Shimojo T, Takakura H;
DR WPI: 1999-080907/07.
DR P-PSDB; AA091836.
XX
PT Recombinant hyperthermostable protease from Pyrococcus furiosus
and gene encoding it, for large scale production of the protease for
industrial use.
XX

```


Claim 6. Page 37-38: 82pp: Japanese.

The invention relates to a hyperthermostable protease derived from a thermophilic bacterium (especially *Pyrococcus furiosus*), the protease has a working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 (optimum 5-8), and retains more than 90% of its activity after 8 hours at 95 deg.C. The invention also provides gene sequences encoding a polypeptide of formula SIG-Ala-Gly-Gly-Asn-Pro, where SIG is a signal peptide from subtilisin, and pro is the above protease. Host cell, (especially *Bacillus* strains) transformed with vectors comprising the genes are used for the recombinant production of the protease. The hyperthermostable protease which can be prepared in quantity suitable for industrial use, can be used as an additive for drugs, washing agents and foodstuffs and for chemical synthesis.

Sequence 1236 BP. 368 A: 251 C: 309 G: 308 T; 0 other;

Query Match 28.3%; Score 560.4; DB 20; Length 1236;
Best Local Similarity 68.0%; Pred. NO. 2.1e-94;
Matches 795; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

Qy	459	CTCTGGCTACGACAGGAACCGCTGTGTGGTCCCATCGTACGGGTATAGACGCCAA	518
Ds	60	CTTGGGATATGATGGTTCTTGGATCAATATAGGAATAATTSACACTGGAAATGACGGTTC	119
Qy	519	CGACCCGATCTGAGGSGCAAGCTCATAGCTTGGTACGACGCCCTCAACGGCGAGTGCAC	578
Ds	120	TCATCCAGATCTCCNAGGMAANGAATTTGGTGGGTAGATTTTGTCAATGGTATGAGGATTA	179
Qy	579	CCCTTACGATACACAGGACACGAGCACCACCTTTCGGGTTATCGTCCCGGAACCGGAG	638
Ds	180	TCCATACGATACCAATGGACATGGAACTCATGTAGCTTCATATAGCACTGGTACTGGAGC	239
Qy	639	CG---ATAACTCCCACTACATAGGCGTCCGCCCGGCGGAGAGCTCGTCGGCGTCMAGGT	695
Ds	240	AGCAAGTAATGCAAGTACAGAGGGAATGGCTCCAGGAGCTAAGCTCGCGGGAATTAAGGT	299
Qy	696	TCCTGGTCGCGACGCTTCGGGAAGCGTCCTCCACCATCATCGCGGTGTTGACTGGGTCGT	755
Ds	300	TCATAGTTCGCGATGGTCTCGAAGCATATCTACTATTAATTAAGGAGGTTGATGGCGCGT	359
Qy	756	CCAGAACAAGGACAAGTACGGGATTAAGGTCATCAACCTCTCCCTCGGCTCTCCACAG	815
Ds	360	TGATAACAAGATAAGTACGGAATTAAGGTCATTATCTTTCTCTGGTTCAGGCCAG	419
Qy	816	CTCCGACGAACCGACTCCCTCAGTCAGGCGGCTCAACAACGCTCGGACGCCGCGTATAGT	875
Ds	420	CTCAGATGGTACTGACGCTTAAGTCAAGGCTGTTAATCGACGCTGGATCTGGATTAGT	479
Qy	876	AGTCTGGTGGCGCGCGCACACGGGCGGAACACTACACCGTCCGGCTACCGCGCAGC	935
Ds	480	TTGTTGGTTGGCCGCTTGGAAACAGTGGACCTACACAGTATACAATCGGTTCTCCAGCA	539
Qy	936	CCCGAGCAAGGTCAATACCGTGGTGAGTTGACACGACACGACACACTCCGAGCTTTC	995
Ds	540	TCCAAGCAAAAGTTATTACAGTTGGAGCGGTGACAAAGTATGATGTTATACAAAGCTCTC	595
Qy	996	CAGCAGGGAGCCGACCGCGACGGAAGGCTCAAGCGGGAAGTCGTCGCCGCCCGCGGTGA	1055
Ds	600	NAGCAGAGGGCCAACTGCACAGCGGAGGCTTAAGCGCTAGAGTTGTGCTCCGGAACAGT	659
Qy	1056	CATCATAGCCCGCGCGCCACGGAACAGGATGGGACCCCGGATAAACAGACTACTACAC	1119
Ds	660	GATAATTGCTGCCNAGAGCAAGTGGAACTAGCATGGGTCAACCAATTAATGACTATTACAC	719
Qy	1116	CAAGGCCCTCGGAACCAAGCATGGCCACCCCGACGTTTCGGCGGTTTGGCGGCTCATCTCT	1175
Ds	720	AGCACTCTCTGGGACATCAATGGCAACTCTCTACGTAAGTGTGATTTGACGCCCTCTGGCT	1179
Qy	1176	CCAGGCCCAACCGAGCTGGACCCCGGACGAAGGTGAAGACCGGCCCTCATCTGAGACCGCGGA	1235
Ds	780	CGAAGCACACCGGAGCTGGAGTCCAGACAAAAGTAAACAGAGCCCTCATGAATGCTGGA	839

Qy	1236	CATAGTCGCGCCCAAGGAGTATACCGGACATCGCCTACGGTGGGGGTAGGCTGAACGCTCTA	12395
Db	840	TATCGTAAAGCCAGATGAATACCGGATATAGCTACGGTCAGGTAGGCTTAATGCGATA	899
Qy	1296	CAAGGCCATCAAGTACGACGCTAGCGCAAGCTCACCTTCACCGCGCTCGCTGGCCGACAA	1355
Db	900	CAAGGCTATTAACCTACGTATATATGCAAGCTAGTGTTCATCTGGATATATGTTGCCAACAA	959
Qy	1356	GGGAAGCGCGACCCACACCTTCAGCGTCAGGGCGCCACCTTCGTGACCGCGCAACCTCTA	1415
Db	960	AGGCAGCGCAATCTACAGTCTGCTATTAGCGGAGCTTGTTCTGTACTCCCACTATA	1019
Qy	1416	CTGGGACACGGGCTCGAGCGCATCGACCTCTACCTCTAGCACCCCAACGGGACAGGCT	1475
Db	1020	CTGGGCAATGGCCAATAGCCAGCTTCATCTTACCTCTACGATGCCAATGGAAACGAGT	1079
Qy	1476	TGACTTACTCTACCGGCTATACGGCTTCGAGAAGTCTGGCTACTACAAACCGCACCGC	1535
Db	1080	TGCTTACTCTTACACGGCTTACATGGATTGGAAGGTTGGTTATTACAAACCCAATGA	1139
Qy	1536	CGGAACCTCGGCGGTCAAGTCTCAGCTCTACAGGGCGGGCGGAACCTTACAGGCTCCAGCT	1595
Db	1140	TGGACATATGGACAATTAAAGTTGTGTAGCTTACAGCGGGAAGTGCAAACTATCACTAGTGT	1199
Qy	1596	CGTACGGCGAGGAGCGCTCAGCCAGTCCGG	1625
Db	1200	GTGAAGTGATGTTCCCTTTTCACAGCCTGG	1229

RESULT 11

RESOL 11
AAT08134
ID AAT08134 standard: DNA: 564 BP.

XX
AC
AST08134-

10-MAY-1996 (first entry);

XX
DE
DNA sequence.

XX processes: hyperthermostable; thermostability; ss.

XX

XX
cc Bureau des finances

XX
NN
W00F365-21

XX
21-DEC-1965

XX
AE - TUN-1005.
95WC-TP01005

XX
00-244-1004
94TB-0173312

PR 13-JUN-1994; 94JP-0130236

PA (TAKI) TAKARA SHUZO CO LTD

PI Asada K, Kato I, Mitla H,

XX

DR P-PSDB; AAR87011.

PT PYROCOCCLUS FURIOSUS HYPER

☰ ☷

XX
22

CC hyperthermostable protease:

CC (AAR87010-11).

SQ Sequence 564 BP; 121 A; 119

Query Match	26.93
-------------	-------

Wed Nov 6 14:29:34 2002

P-PSDB; AB009483.

New DNA sequence of thermophilic protein decomposition enzyme and
protein derived therefrom -

Claim 1; Page 6; 15pp; Korean.

This sequence represents the DNA encoding the thermonarobacter
CC CC yonsei subtilisin-like serine protease of the invention.
XX XX
XX Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 20 other;
SQ

Query Match 9.1%; Score 180.6; DB 23; Length 2121;
Best Local Similarity 55.7%; Pred. No. 2.1e-24;
Matches 395; Conservative 0; Mismatches 299; Indels 15; Gaps 2;

QY 374 TAAAGTTCATACGAGCGATTACAAGTTTAGTTGACGGACGCCTGCCGTCCCAGA 433
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 TAAAGCAAAATGAGTATGATGACGCTGTATACTACTTATAGTACGGCAACAGTGCT 442
QY 434 TAGGGCCGATACCCTCTGGAACCTCCTCGGCTACAGCGAAGCGGTGTGGTGGCCA 493
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 443 TTGGAATCAAAAAGCACCGAGTGATTTTGGAGTACAGAAAAAATAATATACGCAA 502
QY 494 TCGTGATACGGGTATAGACGCGAACACCCCGCATCT---GAAGCGCAAGGTCTATAGCT 550
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 TTATTGACACAGTATAGACGGAATCACGTTGCACCTCTCAGGTGGAAAAAATAATAGSAT 562
QY 551 GTGTACGACGCGGTCAAGCGCAGGTGACGCCCCCTACGATGACACAGGACGAGCAACCCACG 610
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 563 GGNAAGACTTTATCAACACAANAATACACCATACACGACGCAATGGCCATGGNACTCAG 622
QY 611 TTGGGGGTATCGTTTCGGGAACCGGACGAGGTTAATCCCGTACNTAGGCGTGGCCCGCG 670
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 623 TAGCAAGTATTCGTCAGGTACAGGTGCTGGAAACAGTCTTTACAAAGGCCCTTGCTATG 682
QY 671 GCGGGAAGCTCGTCGGCGCTCAAGGTTCTCGGTGCGCAGCGTTTCGGGAAGCGTGTCATCA 730
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 683 ATGCTTTGGTGTAGGATTAAGATTTTAGATGCAATGGAAGCGCAGCATGACGATG 742
QY 731 TCATCGCGGTGTTCACTGGTGTGTCGAGAACAGGACAGTACGGGATAGAGGTCAATCA 790
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 743 TAATCGAGCAATGATGATGGCGTTTCRAAATAAAGATGATATACGGAATCAAAAGTTAT 802
QY 791 ACCTCTCCCTGGCTCCTCCACAGCTCCGACGGAACGACATCCCTCAGTCAGGCGGTC 850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 803 ATTAAAGCCTCGSCACTTCTCAAGTCTGTATGAGTACGACTCTACTCTATTACAGTGA 862
QY 851 ACAAGCCTCTGGGACGCGGCTATAGTASTCTCGGTGGCGCGGCAACAGCGGCGCCGACA 910
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 863 ATAGAGCAGTAGATACGCGTATTTAGTATAGTTTGTAGCAGCAGGAACCTCTGGCCCTG 920
QY 911 CTCTACCGTGGCTCACC CGCGCGCGGACGAGGTCATACCGCTC-----G 950
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 923 ATATACACNTAGAGGTCCCTGGTCTCGGGA AAAAGCCATATACACTCCGCAATGGGAG 982
QY 959 GTSCAGTTGACAGCAAGCAACATCCAGTCTTCCAGCAGGACCGGACCGCGAGG 1018
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 983 ATCTAGGTGAACCTTGGCGTTTAACTCTGCAAGCTTTTCAGCCCGCGGTCTACTGCTGACG 1042
QY 1019 GAAGGCTCARGCGGGAAGTCGTCCGCGCGCGGGCTGTAGATCATATAGCCCC 1067
Db 1043 GAAGATAAAACCTGTACATTCGCGCGCCAGGATATATATACCTGGCGG 1091

Search completed: November 2, 2002, 08:28:13
Job time : 303 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:25:03 ; Search time 30 seconds
(without alignments)
2439.921 Million cell updates/sec

Title: US-09-841-553-5

Sequence: 1 MKGLKALILVLGLVWGS.....YAYSTYCHADQLKAVYVYG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A-Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3428	100.0	659	18	AAW24123
2	2997.5	87.4	654	18	AAW24129
3	2997.5	87.4	654	20	AAW94841
4	2914	85.0	659	18	AAW24121
5	2914	85.0	659	20	AAW94840
6	2349.5	68.5	522	18	AAW24122
7	2349.5	68.5	522	20	AAW94838
8	1918	56.0	412	20	AAW94836
9	1051.5	30.7	237	17	AAW87009
10	864	25.2	168	18	AAW24127
11	862.5	25.2	188	17	AAW87011

12	862.5	25.2	188	18	AAW24126
13	692.5	20.2	734	18	AAW13667
14	692.5	20.2	823	18	AAW13668
15	679.5	19.8	520	18	AAW13666
16	662.5	19.3	1079	22	AAW81180
17	558	16.3	373	11	AAW04137
18	557.5	16.3	373	11	AAW04137
19	557	16.2	379	17	AAW86879
20	556.5	16.2	394	14	AAW86521
21	556	16.2	379	16	AAW26762
22	555	16.2	382	5	AAW08080
23	555	16.2	382	8	AAW04896
24	555	16.2	382	5	AAW04896
25	555	16.2	382	5	AAW04896
26	555	16.2	382	5	AAW04896
27	555	16.2	382	5	AAW04896
28	555	16.2	382	5	AAW04896
29	555	16.2	382	5	AAW04896
30	555	16.2	382	5	AAW04896
31	555	16.2	382	5	AAW04896
32	555	16.2	382	5	AAW04896
33	555	16.2	382	5	AAW04896
34	555	16.2	382	5	AAW04896
35	555	16.2	382	5	AAW04896
36	555	16.2	382	5	AAW04896
37	555	16.2	382	5	AAW04896
38	555	16.2	382	5	AAW04896
39	555	16.2	382	5	AAW04896
40	554.5	16.2	371	17	AAW96242
41	534	16.2	379	19	AAW31600
42	553.5	16.1	373	17	AAW96240
43	553.5	16.1	373	17	AAW96243
44	552.5	16.1	373	17	AAW96241
45	552.5	16.1	373	17	AAW86877

ALIGNMENTS

RESULT 1
AAW24123
ID AAW24123 standard; Protein; 659 AA.
AC AAW24123;
XX
XX
DT 20-APR-1998 (first entry)
XX
DE Protease.
XX
KW Protease; research reagent; thermal stability.
XX
OS Synthetic.
XX
PN W09721823-A1.
XX
PD 19-JUN-1997.
XX
PF 07-NOV-1996; 96WO-JP03253.
XX
PR 12-DEC-1995; 95JP-0303285.
XX
XX (TAKI) TAKARA SHUZO CO LTD
PI Asada K, Kato I, Mitra H, Morishita M, Takakura H;
PI Tsunawawa S, Yamamoto Y.
XX
DR WPI, 1997-332794/30.
XX
PT N-PSDB; AAT85669.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are
PT useful industrially and as research reagents
XX

Thermococcus prote
Streptomyces virid
Dhpa-mal chimeric
Fragment of dhpa g
Transglutaminase f
Sequence of modifi
Subtilisin BPN' mu
Keratinase from Ba
Bacillus carlsberg
Bacillus lichenifo
Sequence of alkali
Bacillus amyloliqu
Amino acid sequenc
Subtilisin from Ba
B. amyloliquetacie
B. amyloliquetacie
B. amyloliquetacie
B. amyloliquetacie
B. amyloliquetacie
Bacillus amyloliqu
B. amyloliquetacie
Mutant subtilisin
Mutant subtilisin
Mutant subtilisin
Mutant subtilisin
Subtilisin BPN' mu

PS Claim 9; Page 92-95; 159pp; Japanese.

XX This sequence is a protease of the invention. The proteases of the
 CC invention have extremely high thermal stability. The proteases can be
 CC used as research reagents, and industrially in the food, drug and
 CC chemical industries.

XX
 XX
 SQ Sequence 659 AA;

Query Match 100.0%; Score 3428; DB 18; Length 659;
 Best Local Similarity 100.0%; Pred. No. 2.3e-209;
 Matches 659; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGLKALILVILVGLVGVSAAPKKEVQVNRVVEKNGYLLPGLFRKIOKLNPEEIS 60
 DB 1 MKGLKALILVILVGLVGVSAAPKKEVQVNRVVEKNGYLLPGLFRKIOKLNPEEIS 60
 QY 61 TVIVFENHREKEIAVRVLELNGAKRVYVYHIIPATAADLKVRDLVLSGLTGKAKLSGV 120
 DB 61 TVIVFENHREKEIAVRVLELNGAKRVYVYHIIPATAADLKVRDLVLSGLTGKAKLSGV 120
 QY 121 RFIQEDYKVTVAELEGDESAAQVMATVYVNLGYDGSIGITIGIDGIDASHPDLOGKV 180
 DB 121 RFIQEDYKVTVAELEGDESAAQVMATVYVNLGYDGSIGITIGIDGIDASHPDLOGKV 180
 QY 181 IGWVDFVNGRSYFYDDHGHGTHVASTAAGTGAASNGKYGKMGAPGAKLAGIKVLGADSGS 240
 DB 181 IGWVDFVNGRSYFYDDHGHGTHVASTAAGTGAASNGKYGKMGAPGAKLAGIKVLGADSGS 240
 QY 241 ISTIIKGVENAVDNKDYGKIKVINLSLGSQSSDGTDSLQAVNNAMDAGIIVCVVAGNS 300
 DB 241 ISTIIKGVENAVDNKDYGKIKVINLSLGSQSSDGTDSLQAVNNAMDAGIIVCVVAGNS 300
 QY 301 GPNTYVGSPPAAASKVITVGVAVDSNDNIASFSSRGPTADGRLEKPEVVAQVDIIAPRASG 360
 DB 301 GPNTYVGSPPAAASKVITVGVAVDSNDNIASFSSRGPTADGRLEKPEVVAQVDIIAPRASG 360
 QY 361 TSMGTPINDIYTKASGTSMAIPHVSGVGLIILQAHPSWTPDKVKTALLETADIVAPKEIA 420
 DB 361 TSMGTPINDIYTKASGTSMAIPHVSGVGLIILQAHPSWTPDKVKTALLETADIVAPKEIA 420
 QY 421 DIAYGAGRVNRYKAIKDYDYAKLTFTGSVADKGSATHTFDVSGATFVTATLWDTGSSDI 480
 DB 421 DIAYGAGRVNRYKAIKDYDYAKLTFTGSVADKGSATHTFDVSGATFVTATLWDTGSSDI 480
 QY 481 DLILYDPNGNEVDYSYATYGFKEKGYVNTAGTWTVKVYVYKGAANYQVDVYSDGLSQ 540
 DB 481 DLILYDPNGNEVDYSYATYGFKEKGYVNTAGTWTVKVYVYKGAANYQVDVYSDGLSQ 540
 QY 541 SGGGNPNPNPNPTTDTQFTFGSVNDYWDTSDFETNNVNSGATKITGDLTDFDYSND 600
 DB 541 SGGGNPNPNPNPTTDTQFTFGSVNDYWDTSDFETNNVNSGATKITGDLTDFDYSND 600
 QY 601 LDILYDPNGNLVDRSTSSNRYEVEANPAGTWTFLYATYSTGNADYQLKAVVYIG 659
 DB 601 LDILYDPNGNLVDRSTSSNRYEVEANPAGTWTFLYATYSTGNADYQLKAVVYIG 659

RESULT 2
 AAW24129
 ID AAW24129 standard; Protein; 654 AA.
 XX
 XX
 AC AAW24129;
 XX
 XX
 DT 20-APR-1998 (first entry)
 XX
 DE Pyrococcus furiosus protease PFUS.
 XX
 KW Protease; research reagent; thermal stability; pyrococcus furiosus
 XX
 OS Pyrococcus furiosus DSM-3638.
 XX
 XX W09721823-A1.

XX 19-JUN-1997.
 XX 07-NOV-1996; 96WO-JP03253.
 XX 12-DEC-1995; 95JP-0323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Asada K., Kato I., Mitsuoka M., Morishita M., Takakura H.,
 PI Tsunashima S., Yamamoto K.
 XX WPI; 1997-332794/30.
 DR N-PSDB; AAT85695.
 XX
 PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are
 PT useful industrially and as research reagents
 PS Disclosure; Page 125-128; 159pp; Japanese.
 XX This sequence represents the protease from Pyrococcus furiosus DSM-3638.
 CC This sequence is a protease of the invention. The proteases of the
 CC invention have extremely high thermal stability. The proteases can be
 CC used as research reagents, and industrially in the food, drug and
 CC chemical industries.

Query Match 87.4%; Score 2997.5; DB 18; Length 654;
 Best Local Similarity 88.3%; Pred. No. 4.7e-182;
 Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MKGLKALILVILVGLVGVSAAPKKEVQVNRVVEKNGYLLPGLFRKIOKLNPEEIS 60
 DB 1 MKGLKALILVILVGLVGVSAAPKKEVQVNRVVEKNGYLLPGLFRKIOKLNPEEIS 60
 QY 61 TVIVFENHREKEIAVRVLELNGAKRVYVYHIIPATAADLKVRDLVLSGLTGKAKLSGV 120
 DB 61 TVIVFENHREKEIAVRVLELNGAKRVYVYHIIPATAADLKVRDLVLSGLTGKAKLSGV 120
 QY 121 RFIQEDYKVTVAELEGDESAAQVMATVYVNLGYDGSIGITIGIDGIDASHPDLOGKV 180
 DB 121 RFIQEDYKVTVAELEGDESAAQVMATVYVNLGYDGSIGITIGIDGIDASHPDLOGKV 180
 QY 181 IGWVDFVNGRSYFYDDHGHGTHVASTAAGTGAASNGKYGKMGAPGAKLAGIKVLGADSGS 240
 DB 181 IGWVDFVNGRSYFYDDHGHGTHVASTAAGTGAASNGKYGKMGAPGAKLAGIKVLGADSGS 240
 QY 241 ISTIIKGVENAVDNKDYGKIKVINLSLGSQSSDGTDSLQAVNNAMDAGIIVCVVAGNS 300
 DB 241 ISTIIKGVENAVDNKDYGKIKVINLSLGSQSSDGTDSLQAVNNAMDAGIIVCVVAGNS 300
 QY 301 GPNTYVGSPPAAASKVITVGVAVDSNDNIASFSSRGPTADGRLEKPEVVAQVDIIAPRASG 360
 DB 301 GPNTYVGSPPAAASKVITVGVAVDSNDNIASFSSRGPTADGRLEKPEVVAQVDIIAPRASG 360
 QY 361 TSMGTPINDIYTKASGTSMAIPHVSGVGLIILQAHPSWTPDKVKTALLETADIVAPKEIA 420
 DB 361 TSMGTPINDIYTKASGTSMAIPHVSGVGLIILQAHPSWTPDKVKTALLETADIVAPKEIA 420
 QY 421 DIAYGAGRVNRYKAIKDYDYAKLTFTGSVADKGSATHTFDVSGATFVTATLWDTGSSDI 480
 DB 421 DIAYGAGRVNRYKAIKDYDYAKLTFTGSVADKGSATHTFDVSGATFVTATLWDTGSSDI 480
 QY 481 DLILYDPNGNEVDYSYATYGFKEKGYVNTAGTWTVKVYVYKGAANYQVDVYSDGLSQ 540
 DB 481 DLILYDPNGNEVDYSYATYGFKEKGYVNTAGTWTVKVYVYKGAANYQVDVYSDGLSQ 540
 QY 541 SGGGNPNPNPNPTTDTQFTFGSVNDYWDTSDFETNNVNSGATKITGDLTDFDYSND 600
 DB 541 SGGGNPNPNPNPTTDTQFTFGSVNDYWDTSDFETNNVNSGATKITGDLTDFDYSND 600
 DB 541 -----PGSSPSPOPEPTVQAK*FQSSDHYHYDRSDTFTFTVNSGATKITGDLTDFDYSND 595

QY 601 LQLYDPNGLVDRSTSSNHYEVEYANPAQGTWFLVAYSTYGMADYQKAVVYVG 659
 DB 596 LQLYDPNGLVDRSTSSNHYEVEYANPAQGTWFLVAYSTYGMADYQKAVVYVG 654

RESULT 3
 ID AAW94841 standard; Protein: 654 AA.
 AC AAW94841;
 DT 06-MAY-1999 (first entry)
 DE Hyperthermostable protease.
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis.
 OS Pyrococcus furiosus.
 PN W09856926-AL.
 PD 17-DEC-1998.
 PF 04-JUN-1998; 98WO-JP02465.
 PR 10-JUN-1997; 97JP-0151969.
 PA (TAKI) TAKARA SHUZO CO LTD.
 PI Asada K, Kato I, Morishita M, Shimojo T, Takakura H;
 PI N-PSDB; AAX05929.
 DR Recombinant hyperthermostable protease from Pyrococcus furiosus -
 PT and gene encoding it, for large scale production of the protease for
 PT industrial use.
 PS Disclosure: Page 60-63; 82pp; Japanese.
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
 CC at 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable
 CC for industrial use, can be used as an additive for drugs, washing agents
 CC and foodstuffs and for chemical synthesis.

QY 601 LQLYDPNGLVDRSTSSNHYEVEYANPAQGTWFLVAYSTYGMADYQKAVVYVG 659
 DB 596 LQLYDPNGLVDRSTSSNHYEVEYANPAQGTWFLVAYSTYGMADYQKAVVYVG 654

Query Match 87.4%; Score 2997.5; DB 20; Length 654;
 Best Local Similarity 88.3%; Pred. No. 4,7e-182;
 Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MKGLKALLIVLVLGLVGVSAAPKKEQVRYNVEKNYGLLTPGLPRKTKLNPNEETS 60
 DB 1 MKGLKALLIVLVLGLVGVSAAPKKEQVRYNVEKNYGLLTPGLPRKTKLNPNEETS 60
 QY 61 TVIVFNHREKIAVRVLELNGAKRVYTHIIPAIAADLKVRDLVTSGLTGKAKLSGV 120
 DB 61 TVIVFNHREKIAVRVLELNGAKRVYTHIIPAIAADLKVRDLVTSGLTGKAKLSGV 120
 QY 121 RFIQEDKYVTSAELEGLDESAQAQVATYWNILGDSGCIIGIITDIDASHPDQKGV 180
 DB 121 RFIQEDKYVTSAELEGLDESAQAQVATYWNILGDSGCIIGIITDIDASHPDQKGV 180
 QY 181 ICWDFVNGRSYPYDDHGHGTHVASTAAGTGAASNGYKGNAPGAKLAGIKVLGADGSGS 240

DB 181 ICWDFVNGRSYPYDDHGHGTHVASTAAGTGAASNGYKGNAPGAKLAGIKVLGADGSGS 240
 QY 241 ISTIKGEVAVDNKDKALIKVINLSGSSQSDGTDLSQAVNNAWDACIVVCVAAGNS 300
 DB 241 ISTIKGEVAVDNKDKYGIKVINLSGSSQSDGTDLSQAVNNAWDACIVVCVAAGNS 300
 QY 301 GPNYTVGSPAASAKVITYGAVDSNDNIASPSRGPTADGRKLPVAVPAGVDTIAPRASG 360
 DB 301 GPNYTVGSPAASAKVITYGAVDSNDNIASPSRGPTADGRKLPVAVPAGVDTIAPRASG 360
 QY 361 TSMGTPINDYTYKASCTSMATPHVSGVGCALILQAHPSWTPDKVKYKTALITETADIAPKEIA 420
 DB 361 TSMGTPINDYTYKASCTSMATPHVSGVGCALILQAHPSWTPDKVKYKTALITETADIAPKEIA 420
 QY 421 DIAYGAGRVNAYKAIYDPAKLTFTGVSADKGSATHTFDVSGATFVATLVWDTGSSDI 480
 DB 421 DIAYGAGRVNAYKAIYDPAKLTFTGVSADKGSATHTFDVSGATFVATLVWDTGSSDI 480
 QY 481 DLXLYDPNGNEVDYSYATYGFPEKVGYNPTAGTWTWVKVYSGAANYQDVWSDGSLSQ 540
 DB 481 DLXLYDPNGNEVDYSYATYGFPEKVGYNPTAGTWTWVKVYSGAANYQDVWSDGSLSQ 540
 QY 541 SGGNPNPNPNPTPTTDTQTFTGSVNDYWDTSFTTMVNSGATKITGDLTDTSYND 600
 DB 541 SGGNPNPNPNPTPTTDTQTFTGSVNDYWDTSFTTMVNSGATKITGDLTDTSYND 600
 QY 601 LDLYLDPNGNLDVSTSSNHYEVEYANPAQGTWFLVAYSTYGMADYQKAVVYVG 659
 DB 596 LDLYLDPNGNLDVSTSSNHYEVEYANPAQGTWFLVAYSTYGMADYQKAVVYVG 654

RESULT 4
 ID AAW24121 standard; Protein: 659 AA.
 AC AAW24121;
 DT 20-APR-1998 (first entry)
 DE Thermococcus protease.
 KW Protease; research reagent; thermal stability; thermococcus celer.
 OS Thermococcus celer DSM-2476.
 PN W09721823-AL.
 PD 19-JUN-1997.
 PF 07-NOV-1996; 96WO-JP01253.
 PR 12-DEC-1995; 95JP-032245.
 PA (TAKI) TAKARA SHUZO CO LTD.
 PI Asada K, Kato I, Mitsu M, Morishita M, Takakura H;
 PI Tsunashima S, Yamamoto K;
 PI WPI; 1997-332794/30.
 PI N-PSDB; AAT85667.

Protease(s) and genes encoding them obtained from Thermococcus and
 Pyrococcus strains - have extremely high thermal stability and are
 useful industrially and as research reagents
 Claim 1; Page 82-85; 159pp; Japanese.
 This sequence represents the protease from Thermococcus celer DSM-2476.
 This sequence is a protease of the invention. The proteases of the
 invention have extremely high thermal stability. The proteases can be
 used as research reagents, and industrially in the food, drug and
 chemical industries.

QY 533 VSDGSLSQSGGPNPNPNTPTTDTQFTGVSNDYWDTSFTMNVNSGATKITGDL 592
 DQ 533 VSDGSLSQSGGPNPNPNTPTTDTQFTGVSNDYWDTSFTMNVNSGATKITGDL 592
 QY 593 TFDTSNDLDLYDPGNDLDRSTSSNYSYEHVEYANPAGTWTPLVYAYSTYGNADYOL 652
 DQ 593 TFDTSNDLDLYDPGNDLDRSTSSNYSYEHVEYANPAGTWTPLVYAYSTYGNADYOL 652
 QY 653 KAVVYTG 659
 DQ 653 KAVVYTG 659

RESULT 6
 AAW24122 standard; Protein: 522 AA.
 AC AAW24122:
 DT 20-APR-1998 (first entry)
 XX Pyrococcus furiosus protease.
 XX Protease; research reagent; thermal stability; pyrococcus furiosus
 XX Pyrococcus furiosus DSM-3638.
 OS Key Location/Qualifiers
 FH Misc-difference 428
 FT /label= Gly, Val
 FT /note= "encoded by GNA"
 XX
 XX HQ9721823-A1.
 XX 19-JUN-1997.
 XX 07-NOV-1996; 98WO-JP03253.
 XX 12-DEC-1995; 95JP-0323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Asada K, Kato I, Mitu M, Morishita M, Takakura H;
 PI Tsunashima S, Yamamoto K;
 XX MPI: 1997-332794/30.
 XX N-PSDB; AAT85668.
 XX
 XX Protease(s) and genes encoding them obtained from Thermococcus and;
 XX Pyrococcus strains - have extremely high thermal stability and are;
 XX useful industrially and as research reagents
 XX Claim 5: Page 87-90; 159pp; Japanese.
 XX This sequence represents the protease from Pyrococcus furiosus DSM-3638.
 XX This sequence is a protease of the invention. The proteases of the
 XX invention have extremely high thermal stability. The proteases can be
 XX used as research reagents, and industrially in the food, drug and
 XX chemical industries.
 XX Sequence 522 AA;

Query Match 68.5%; Score 2349.5; DB 18; Length 522;
 Best Local Similarity 85.2%; Pred. No. 4.5e-141;
 Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELEGDESAAQVATYNNLGYDGSIGITIGIDTGDASHPDLOGKVICWDFVNGRSY 192
 DQ 1 AELEGDESAAQVATYNNLGYDGSIGITIGIDTGDASHPDLOGKVICWDFVNGRSY 60
 QY 193 PYDDHGHGHTVASTAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTITIKGVMAV 522
 DQ 193 PYDDHGHGHTVASTAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTITIKGVMAV 522

DB 61 PYDDHGHGHTVASTAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTITIKGVMAV 120
 QY 253 DNKDXGKIKVINLSLGSQSSDGTDSLSQAVNNANDGICVCAAGNSGPNPTVYVGSAA 312
 DQ 121 DNKDXGKIKVINLSLGSQSSDGTDSLSQAVNNANDGICVCAAGNSGPNPTVYVGSAA 180
 QY 313 ASKYTVVGVADVNDONIASPSSRCPTADGRKPEVAPGVDIIPRASGTSMTGTPINDYTT 372
 DQ 181 ASKYTVVGVADVNDONIASPSSRCPTADGRKPEVAPGVDIIPRASGTSMTGTPINDYTT 240
 QY 373 KASGTSNATPHVSGVGLLQAHPSMTDKVKTALLETADIAPKEIADIAGAGRVNY 432
 DQ 241 KASGTSNATPHVSGVGLLQAHPSMTDKVKTALLETADIAPKEIADIAGAGRVNY 300
 QY 433 KAIKDYDIKLTFTGVSADKGSATHTFVSGATFVTATLYMDTSGSDIDLVDLPNGNEV 492
 DQ 301 KAIKDYDIKLTFTGVSADKGSATHTFVSGATFVTATLYMDTSGSDIDLVDLPNGNEV 360
 QY 493 DYSYATYGFVKVYNNPTAGTWTWKVYKGAANYOVDSGSLSSQSGGPNPNP 552
 DQ 361 DYSYATYGFVKVYNNPTAGTWTWKVYKGAANYOVDSGSLSSQSGGPNPNP 415
 QY 553 NPTPTDTDTGVSNDYWDTSFTMNVNSGATKITGDLTFTSYNDLDLYDPNGNL 612
 DQ 416 NPTPTDTDTGVSNDYWDTSFTMNVNSGATKITGDLTFTSYNDLDLYDPNGNL 475
 QY 613 VDRSTSSNYSYEHVEYANPAGTWTPLVYAYSTYGNADYOLKAVVYTG 659
 DQ 476 VDRSTSSNYSYEHVEYANPAGTWTPLVYAYSTYGNADYOLKAVVYTG 522

RESULT 7
 AAW94838 standard; Protein: 522 AA.
 AC AAW94838:
 DT 06-MAY-1999 (first entry)
 XX Hyperthermostable protease.
 XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX additive; drug; washing agent; foodstuff; chemical synthesis.
 XX Pyrococcus furiosus.
 XX Key Location/Qualifiers
 FH Misc-difference 428
 FT /label= Gly or Val
 XX WO9856926-A1.
 XX 17-DEC-1998.
 XX 04-JUN-1998; 98WO-JP02465.
 XX 10-JUN-1997; 97JP-0151919.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Asada K, Kato I, Morishita M, Shimojo T, Takakura H;
 XX MPI: 1999-080907/07.
 XX Recombinant hyperthermostable protease from Pyrococcus furiosus -
 XX and gene encoding it, for large scale production of the protease for
 XX industrial use.
 XX Claim 1: Page 39-41; 82pp; Japanese.
 XX The invention relates to a hyperthermostable protease derived from a
 XX thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 XX working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10

CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
 CC at 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable
 CC for industrial use, can be used as an additive for drugs, washing agents
 CC and foodstuffs and for chemical synthesis.

XX Sequence 522 AA;

Query Match 58.5%; Score 2349.5; DB 20; Length 522;
 Best Local Similarity 85.2%; Pred. No. 4.5e-141;
 Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELGLDESAAOVNATYVNNLGYDSGGITIGITDGTGASHDPOGKYGIVGVDFVNGRSY 192
 DB 1 AELGLDESAAOVNATYVNNLGYDSGGITIGITDGTGASHDPOGKYGIVGVDFVNGRSY 60
 QY 193 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAY 252
 DB 61 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAY 120
 QY 253 DNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNAWDAGIWCVAAGNSGPNNTYVGSPPAA 312
 DB 121 DNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNAWDAGIWCVAAGNSGPNNTYVGSPPAA 180
 QY 313 ASKVTITVGVDSNDNTAFSSRCPTADGRKLPKPEVAPGVDDIAPRASGTSMTGPIINDYIT 372
 DB 181 ASKVTITVGVDSNDNTAFSSRCPTADGRKLPKPEVAPGVDDIAPRASGTSMTGPIINDYIT 240
 QY 373 KASGTSNATPHVSGVAGLILQAHPSWTPDKVKTALITETADIIVAPKEIADIAGAGRVNAY 432
 DB 241 AAPGTSNATPHVAGIAAALLQAHPSWTPDKVKTALITETADIIVAPKEIADIAGAGRVNAY 300
 QY 433 KAIKYDDYAKLTFTGSAVKAGSANTHTFDVSGATFVATLYWDTGSSDLDLYDPNGEV 492
 DB 301 KAIKYDDYAKLTFTGSAVKAGSANTHTFDVSGATFVATLYWDTGSSDLDLYDPNGEV 360
 QY 493 DYSYATYIGFEKVGYYNPAGTWKVVYKGAANYQVVDVSDGSLSGGNGPNPNP 552
 DB 361 DYSYATYIGFEKVGYYNPAGTWKVVYKGAANYQVVDVSDGSLSGGNGPNPNP 415
 QY 553 NPTPTDPTQFTGSDVNDYDTSDFTNMNSGATKITGDLTFTSYNDLDLYDPNGEV 612
 DB 416 QPETHVDAKTFQSDHYIYDRSDFTWVNSGATKITGDLTFTSYNDLDLYDPNGEV 475
 QY 613 VDRSSNSYEHVEVYANPAGTWKVVYKGAANYQVVDVSDGSLSGGNGPNPNP 659
 DB 476 VDRSSNSYEHVEVYANPAGTWKVVYKGAANYQVVDVSDGSLSGGNGPNPNP 522

RESULT 8

ID AAW94836 standard; Protein; 412 AA.

XX AAW94836;

XX 06-MAY-1999 (first entry)

DE Hyperthermostable protease fragment.

KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;

KW additive; drug; washing agent; foodstuff; chemical synthesis.

XX Pyrococcus furiosus.

OS WO9856926-A1.

PN 17-DEC-1998.

PD 04-JUN-1998; 98WO-JP02465.

XX 10-JUN-1997; 97JP-0151569.
 PR (TAKI) TAKARA SHUZO CO LTD.
 PA Asada K, Kato I, Morishita M, Shimojo T, Takakura H;
 PI WPI: 1999-080907/07.
 DR N-PSDB; AAX05920.
 XX Recombinant hyperthermostable protease from Pyrococcus furiosus -
 PT and gene encoding it, for large scale production of the protease for
 PT industrial use.

PS Claim 2; Page 35-37; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
 CC at 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable
 CC for industrial use, can be used as an additive for drugs, washing agents
 CC and foodstuffs and for chemical synthesis.

XX Sequence 412 AA;

Query Match 56.0%; Score 1918; DB 20; Length 412;
 Best Local Similarity 89.5%; Pred. No. 7.8e-114;
 Matches 367; Conservative 15; Mismatches 28; Indels 0; Gaps 0;

QY 133 AELGLDESAAOVNATYVNNLGYDSGGITIGITDGTGASHDPOGKYGIVGVDFVNGRSY 192
 DB 1 AELGLDESAAOVNATYVNNLGYDSGGITIGITDGTGASHDPOGKYGIVGVDFVNGRSY 60
 QY 193 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAY 252
 DB 61 PYDDHGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAY 120
 QY 253 DNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNAWDAGIWCVAAGNSGPNNTYVGSPPAA 312
 DB 121 DNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNAWDAGIWCVAAGNSGPNNTYVGSPPAA 180
 QY 313 ASKVTITVGVDSNDNTAFSSRCPTADGRKLPKPEVAPGVDDIAPRASGTSMTGPIINDYIT 372
 DB 181 ASKVTITVGVDSNDNTAFSSRCPTADGRKLPKPEVAPGVDDIAPRASGTSMTGPIINDYIT 240
 QY 373 KASGTSNATPHVSGVAGLILQAHPSWTPDKVKTALITETADIIVAPKEIADIAGAGRVNAY 432
 DB 241 AAPGTSNATPHVAGIAAALLQAHPSWTPDKVKTALITETADIIVAPKEIADIAGAGRVNAY 300
 QY 433 KAIKYDDYAKLTFTGSAVKAGSANTHTFDVSGATFVATLYWDTGSSDLDLYDPNGEV 492
 DB 301 KAIKYDDYAKLTFTGSAVKAGSANTHTFDVSGATFVATLYWDTGSSDLDLYDPNGEV 360
 QY 493 DYSYATYIGFEKVGYYNPAGTWKVVYKGAANYQVVDVSDGSLSGGNGPNPNP 542
 DB 361 DYSYATYIGFEKVGYYNPAGTWKVVYKGAANYQVVDVSDGSLSGGNGPNPNP 410

RESULT 9

AA87009 standard; Proteins; 237 AA.

XX AA87009;

XX 10-MAY-1996 (first entry)

XX Hyperthermostable protease.

XX Asada K, Kato I, Mitta M, Morishita M, Tsunasawa S;
 PI Yamamoto K;
 XX WPI; 1996-049674/05.
 DR Pyrococcus furiosus hyperthermostable protease gene - useful for
 PT recombinant prodn. of hyperthermostable protease
 XX Example 4; Page 66-67; 85pp; Japanese.
 PS The invention relates to a gene (AAT08141) that codes for a
 CC hyperthermostable protease (AAR97007) of Pyrococcus furiosus.
 CC 2 DNA sequences (AAT08133-34) are provided encoding peptides
 CC (AAR87010-11).
 XX Sequence 188 AA;
 SO Query Match 25.2%; Score 862.5; DB 17; Length 188;
 Best Local Similarity 89.4%; Pred. NO. 2.6e-47;
 Matches 168; Conservative 8; Mismatches 11; Indels 1; Gaps 1;
 QY 199 HGTHVASIAGTGAASNGKYGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKY 258
 DB 2 HGTHVAGTGTGVSV-NSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVQNKDKY 60
 QY 259 GIKVINLSGSSQSSDGTDSLQAVNNADAGIWCVAAGNSGPNNTYTVGSPAASKVIT 318
 DB 61 GIRVINLSGSSQSSDGDLSQAVNNADAGIWCVAAGNSGPNNTYTVGSPAASKVIT 120
 QY 319 VGAVDSNDNIASFSSRGPTADGRKLPVWAPGVDDIAPRASGTSMTGTPINDYITKASGTS 378
 DB 121 VGAVDSNDNIASFSSRGPTADGRKLPVWAPGVDDIAPRASGTSMTGTPINDYITKASGTS 180
 QY 379 MATPHVSG 386
 DB 181 MATPHVTG 188
 RESULT 12
 AAW24126
 TD AAW24126 standard; Protein: 188 AA.
 XX AC AAW24126;
 XX 20-APR-1998 (first entry)
 DE Thermococcus protease fragment.
 KW Protease; research reagent; thermal stability: thermococcus celer.
 XX OS Thermococcus celer DSM-2476.
 XX PN WO9721823-A1.
 XX PD 19-JUN-1997.
 XX PF 07-NOV-1996; 96WO-JP03253.
 XX PR 12-DEC-1995; 95JP-0323285.
 XX PA (TAKI) TAKARA SHUZO CO LTD.
 XX PI Asada K, Kato I, Mitta M, Morishita M, Takakura H;
 PI Tsunasawa S, Yamamoto K;
 DR WPI; 1997-332794/30.
 DR N-PSDB; AAT85676.
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are
 PT useful industrially and as research reagents

PS Disclosure: Page 110-112; 159pp; Japanese.
 CC This sequence represents a fragment of the protease from Thermococcus
 CC celer DSM-2476 (see AAW24121 for full length sequence). This sequence is
 CC a fragment of the protease of the invention. The proteases of the
 CC invention have extremely high thermal stability. The proteases can be
 CC used as research reagents, and industrially in the food, drug and
 CC chemical industries.
 XX Sequence 188 AA;
 SO Query Match 25.2%; Score 862.5; DB 18; Length 188;
 Best Local Similarity 89.4%; Pred. NO. 2.6e-47;
 Matches 168; Conservative 8; Mismatches 11; Indels 1; Gaps 1;
 QY 199 HGTHVASIAGTGAASNGKYGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKY 258
 DB 2 HGTHVAGTGTGVSV-NSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDMVQNKDKY 60
 QY 259 GIKVINLSGSSQSSDGTDSLQAVNNADAGIWCVAAGNSGPNNTYTVGSPAASKVIT 318
 DB 61 GIRVINLSGSSQSSDGDLSQAVNNADAGIWCVAAGNSGPNNTYTVGSPAASKVIT 120
 QY 319 VGAVDSNDNIASFSSRGPTADGRKLPVWAPGVDDIAPRASGTSMTGTPINDYITKASGTS 378
 DB 121 VGAVDSNDNIASFSSRGPTADGRKLPVWAPGVDDIAPRASGTSMTGTPINDYITKASGTS 180
 QY 379 MATPHVSG 386
 DB 181 MATPHVTG 188
 RESULT 13
 AAW13667
 ID AAW13667 standard; Protein: 734 AA.
 XX AC AAW13667;
 XX 06-OCT-1997 (first entry)
 DE Streptomyces viridosporus dhpA gene product.
 KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease.
 XX OS Streptomyces viridosporus.
 XX FH Key Location/Qualifiers
 FT Peptide 205..724
 FT /note="fragment of dhpA protein; see AAW13666"
 XX PN WO9705243-A1.
 XX PD 13-FEB-1997.
 XX PF 30-JUL-1996; 96WO-JP03147.
 XX PR 29-FEB-1996; 96JP-0067478.
 XX PR 31-JUL-1995; 95JP-0212975.
 XX PA (SAOC) MERCIAN CORP.
 XX PI Arisawa A, Dobashi K, Teshiki K, Matsufuji M, Nakashima T;
 PI Tsuruta T, Yoshiooka T;
 XX WPI; 1997-145682/13.
 DR N-PSDB; AAT61454.
 XX Asymmetric hydrolase gene derived from Streptomyces viridosporus -
 PT acts on 4-substituted-1,4-di-hydro-pyridine derivatives to produce
 PT chiral derivatives useful for synthesis of cardiovascular drugs

PS Claim 3; Page 49-55; 78pp; Japanese.

XX This sequence is an asymmetric hydrolase encoded by the Streptomyces
CC viridosporus dhpA gene. The enzyme acts on 4-substituted-1,4-
CC dihydropyridine derivatives. The enzyme allows the efficient conversion
CC of 4-substituted-1,4-dihydropyridine esters to chiral partially
CC hydrolysed derivatives, for use in the synthesis of cardiovascular drugs
CC suitable for the treatment of e.g. hypertension and ischaemic heart
CC disease.

XX SQ Sequence 734 AA;

Query Match 20.2%; Score 692.5; DB 18; Length 734;
Best Local Similarity 33.6%; Pred. No. 9.1e-36;
Matches 202; Conservative 62; Mismatches 223; Indels 115; Gaps 21;

OY 110 LTGGKAKLSGVRFQEDYKVTYSAELEGDESAQAQVMAVYVNLGVGSGITIGIDTGI 169
DB 182 VTNGDRTASGIARHVLDD-----GVRRAALDTSVGQICAPKASAGYDGKVIADVLTGV 236
OY 170 DASHPDLOGKVIQWDFVNGRSYPYDDHGHGTHVASTAGTGAASNGYKGMAPGAKLAG 229
DB 237 DTSHPDLKGRVTAASKNFTAPG-AGDKVGHGTHVASTAGTGAOSKGYKGVAPCAATLN 295
OY 230 IKVLGADCGSGISTTIKGYEAVDNKDKYIKVINLSLGSQSSDGTDSLQAVNN-AMD 288
DB 296 GKVLDDSGFGDSDGILLAGMEWAA-----AQGADVYVNSLGGMDTPE-TDPLEAAVDKLSAE 350
OY 289 AGIVVCVAAGNSGPNVTYVGSAAASKVITVGAVDSDNDNIASFSSRGP-TADGRUKPEVY 347
DB 351 KGVLFALTAAGNEGPE--SIGSPGSADAALTVGAVDDKDLADESTGPRUGDGAIKPDVT 408
OY 348 APGVDDIAPRASGTSNGTPIIND---YTKASGTSMTATPHVSGVAGLILQAHPSWTPDKVK 404
DB 526 LTYRNLGTQDVTLLKLTSTADPKGKAPAGFTLGLATTIVPA---CGSASVDWNTADTKL 582
OY 489 GNEYDYSTAYYGFERYKYNRTAGTWYKVVSY-----KGAANYQVDDVSDGS 537
DB 583 GCTVDCATSAVWATGGQVRTAAAVOREVESYDVTVRHIGRDKGKPTTEHLDLIGYAG 642
OY 538 LSGSGSGNPNPNPNPTTDTOTF---TGS-VNDYWDTSDTFM-----WVNSQA 585
DB 643 LSGSGRYG-----APATDTATLRLPKGYLVDSIAKDFGLKGGIDMLVQPKLSV 693
OY 586 TKITGDLTFTD-SYNDLDLYDPNGNLDRSTSSNYSHEVYANPAGPTWTELVVAYST 644
DB 694 TRDT-TLTLDAARTTKAADITVPDPK-----AKPLSAT---IGTYDT 731
OY 645 YG 646
DB 732 AG 733

RESULT 14

AAW13668
ID AAW13668 standard; Protein; 823 AA.

XX AC AAW13668;

XX DT 06-OCT-1997 (first entry)

XX DB DhpA-mel chimeric protein.

XX KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;

XX derivative; Streptomyces viridosporus; ester; chiral; synthesis;

KW cardiovascular; treatment; hypertension; ischaemic heart disease.

XX Chimeric Streptomyces viridosporus;
OS Chimeric Streptomyces antibioticus.

XX FH Key Location/Qualifiers
FT Protein 1..734
FT Peptide /label= DhpA_protein_product
FT 205..724
FT /note= "see AAW13666"
FT Peptide 735..823
FT /note= "melanin (partial sequence)"

XX W09705243-A1.

XX PN 13-FEB-1997.

XX PD 30-JUL-1996; 96WO-JP02147.

XX PF 29-FEB-1996; 96JP-0067478.

XX PR 31-JUL-1995; 95JP-0212975.

XX PA (SAOC) MERCIAN CORP.

XX PI Arisawa A, Dobaashi K, Ishihara K, Matsufuji M, Nakashima T;

PI Tsuruta T, Yoshida K;

XX DR MPI, 1997-145682/13.

XX DR N-PSDB; AAT61455.

XX PT Asymmetric hydrolase gene derived from Streptomyces viridosporus -
PT acts on 4-substituted-1,4-dihydropyridine derivatives to produce
PT chiral derivatives useful for synthesis of cardiovascular drugs

XX PS Claim 5; Page 37-43; 78pp; Japanese.

XX This sequence is a fusion protein comprising Streptomyces viridosporus
CC dhpA gene product (an asymmetric hydrolase) which acts on 4-substituted-
CC 1,4-dihydropyridine derivatives, and melanin from S. antibioticus. The
CC DhpA enzyme allows the efficient conversion of 4-substituted-1,4-
CC dihydropyridine esters to chiral partially hydrolysed derivatives,
CC for use in the synthesis of cardiovascular drugs suitable for the
CC treatment of e.g. hypertension and ischaemic heart disease.

XX SQ Sequence 823 AA;

Query Match 20.2%; Score 692.5; DB 18; Length 823;
Best Local Similarity 33.6%; Pred. No. 1.1e-35;
Matches 202; Conservative 62; Mismatches 223; Indels 115; Gaps 21;

OY 110 LTGGKAKLSGVRFQEDYKVTYSAELEGDESAQAQVMAVYVNLGVGSGITIGIDTGI 169
DB 182 VTNGDRTASGIARHVLDD-----GVRRAALDTSVGQICAPKASAGYDGKVIADVLTGV 236
OY 170 DASHPDLOGKVIQWDFVNGRSYPYDDHGHGTHVASTAGTGAASNGYKGMAPGAKLAG 229
DB 237 DTSHPDLKGRVTAASKNFTAPG-AGDKVGHGTHVASTAGTGAOSKGYKGVAPCAATLN 295
OY 230 IKVLGADCGSGISTTIKGYEAVDNKDKYIKVINLSLGSQSSDGTDSLQAVNN-AMD 288
DB 296 GKVLDDSGFGDSDGILLAGMEWAA-----AQGADVYVNSLGGMDTPE-TDPLEAAVDKLSAE 350
OY 289 AGIVVCVAAGNSGPNVTYVGSAAASKVITVGAVDSDNDNIASFSSRGP-TADGRUKPEVY 347
DB 351 KGVLFALTAAGNEGPE--SIGSPGSADAALTVGAVDDKDLADESTGPRUGDGAIKPDVT 408
OY 348 APGVDDIAPRASGTSNGTPIIND---YTKASGTSMTATPHVSGVAGLILQAHPSWTPDKVK 404
DB 409 APGVDDIAPRASGTSNGTPIIND---YTKASGTSMTATPHVSGVAGLILQAHPSWTPDKVK 468
OY 405 TALLETADIVAPKETADIAYGAGRVNYKAIK-----AKPLSAT---IGTYDT 436
DB 469 GAL--TGSTKGGK-YTPEGSGRLOADKALQOTVITADPVSVSGVQOHPHTDDEPVTKO 525

```

CC suitable for the treatment of e.g. hypertension and ischaemic heart
XX disease.
SQ Sequence 520 AA;

Query Match 19.8% Score 679.5; DB 18; Length 520;
Best Local Similarity 36.3%; Pred. No. 3.9e-35;
Matches 181; Conservative 55; Mismatches 183; Indels 79; Gaps

QY 138 LDSEAAQVNATVWNLGYSQSGTITIGITDIDASHDLGGKVGWGVDFNGRSYPYDDH 197
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 LDVSVGIGAPKANSAGYDGGKYIAVLDTGVDSHPDLAKRVTASKNFTAAPG-AGDKV 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 198 GHGTHVASTAAGTGAASNGKYNAPACAKLAGIKVLGADSGSISTIIKGVHAVDKNK 257
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 GHGTHVASTAAGTGAQSGKYGKVGAPGAAILNGKVLDDSGFGDGGILLACHEWAA----A 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 258 YGKIVNLGLSGSSOSSDGTDLSCAYNN-AMDAIVYVCAAGNSGPNYTYVGSPPAAASKV 316
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 115 QGADVYVNMSLGCMOTPE--TDPLSAAVGLSAAKEGVLFIAAGNEGPE--SIGSPGSADAA 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 317 ITVCAYDSNDNIASFSSRGF-TADGRLEPEVAVGVDIAPRASGTSMGTPIND---YKT 372
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 173 LTVGAYDDDKDLADESTGPRGLDGAIRPDTATGVDDITPAASRENDIQEVEGEPAGYW 232
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 373 KASGTSMATPHVSGVGLIILZAIHPSWTPDKVKTLITADITVAPREIADIAYGAGRVVY 432
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 233 TISGTSMATPHVAGAAALCKQHPDWSAELKGLAL--TGSTKGGR-YTFPGSGSGRIQAD 289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 433 KAIK-----YDDYAKLFTIGSVADKGSAT 456
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 290 KALQQTIVADPVSVSFGVOQWPHIDDEPVTQLTYRNLGTQDVTLLKLTATDPKGAAP 349
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 457 HTFDVSGATFEVATLXWDTGSSDIDLXLYDNGNEVDYSVTAYFEKVGYYNPACTWT 516
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 AGFFTIGATTVTVA---GGASVDMATADRLGGTVDGAYSAVYVATCGGTVETAAVQ 406
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 517 VKVSV-----KGACQVQDVVYSDGSLSQSGGHNPNENPNPTTDTQTF-- 563
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 407 REVESIDVTYRHIGDCKPTTEHLTDLIGYAGLGSGRGV-----APATDTATLRL 457
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 564 -TGS-VNDYWDTSOTPM 579
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 458 PKGTLYLVDWSIAKDFGL 475
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: October 31, 2002, 13:29:43
Job time : 33 secs

```

Search completed: October 31, 2002, 13:29:43
Job time : 33 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:28:38 ; Search time 13 Seconds

(without alignments)
1238.190 Million cell updates/sec

Title: US-09-841-553-5

Perfect score: 3428

Sequence: 1 MGLKALLVILVGLVGS.....YASTYGNADYQLKAVYVG 659

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
 - 2: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
 - 3: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
 - 4: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
 - 5: /cgn2_6/ptodata/1/1aa/PTUS-COMB.pep:*
 - 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3428	100.0	659	4	US-08-894-818B-5
2	2997.5	87.4	654	4	US-08-894-818B-35
3	2997.5	87.4	654	4	US-09-445-472-15
4	2914	85.0	659	4	US-08-894-818B-1
5	2914	85.0	659	4	US-09-445-472-12
6	2349.5	68.5	522	4	US-08-894-818B-3
7	2349.5	68.5	522	4	US-09-445-472-4
8	1948	56.0	412	4	US-09-445-472-1
9	1045.5	30.5	237	1	US-08-750-532-18
10	692.5	20.2	734	4	US-09-000-016-4
11	692.5	20.2	823	4	US-09-000-016-2
12	679.5	19.8	520	4	US-09-000-016-7
13	557	16.2	379	1	US-08-685-774-2
14	555	16.2	382	4	US-09-445-472-31
15	554.5	16.2	378	1	US-07-772-087-A
16	534	16.2	379	2	US-08-845-295A-1
17	534	16.2	379	3	US-09-140-933-1
18	534	16.2	379	4	US-09-146-661-1
19	534	16.2	379	4	US-09-150-515-1
20	551	16.1	382	4	US-09-255-502-2
21	550	16.0	382	6	5472855-2
22	549.5	16.0	381	1	US-07-772-087-6
23	549.5	16.0	381	1	US-08-173-508-12
24	549.5	16.0	381	2	US-08-265-310-12
25	549.5	16.0	381	3	US-08-951-742-12
26	547	16.0	382	4	US-09-178-155-2
27	545	15.9	350	1	US-07-923-260A-4

28	543.5	15.9	352	1	US-07-923-260A-2
29	543	15.8	383	2	US-08-460-343B-2
30	543	15.8	392	1	US-08-398-028B-2
31	543	15.8	302	2	US-08-504-265B-2
32	540.5	15.8	1398	1	US-08-750-532-9
33	540.5	15.8	1398	1	US-08-894-818B-8
34	540.5	15.8	1398	4	US-09-445-472-6
35	537.5	15.7	380	1	US-07-518-318-2
36	537.5	15.7	380	2	US-08-413-724-2
37	537.5	15.7	380	3	US-09-049-867-2
38	537.5	15.7	380	4	US-08-853-494-2
39	537.5	15.7	380	4	US-08-716-293-1
40	537.5	15.7	380	4	US-08-269-050-2
41	537.5	15.7	380	6	5217878-2
42	537.5	15.7	380	6	5336611-2
43	536	15.6	377	1	US-07-772-087-2
44	535	15.6	382	7	US-08-504-265B-75
45	531	15.5	274	4	US-09-104-623A-3

ALIGNMENTS

RESULT 1
US-08-894-818B-5
Sequence 5, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, HIKARU
APPLICANT: MORISHITA, MIO
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Nelmark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 332285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger
REGISTRATION NUMBER: 35,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5197
TELEFAX: (202) 737-3534
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 659 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 6, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Patent No. 5217878
Patent No. 5336611
Sequence 2, Appl
Sequence 3, Appl

US-08-894-818B-5

Query Match 100.08; Score 3428; DB 4; Length 659;
Best Local Similarity 100.08; Pred. No. 1.1e-241; Indels 0; Gaps 0;
Matches 659; Conservative 0; Mismatches 0;

QY 1 MKGLKALILVILVGLVGSVAAAPKKEQVNRVKNKGLTGLPKFKIQLKPNNEIS 60
DB 1 MKGLKALILVILVGLVGSVAAAPKKEQVNRVKNKGLTGLPKFKIQLKPNNEIS 60

QY 61 TVIVFENHREKEJAVRVLKMGAKRVYVYHIIPAIADLKVRDLVLISGLTGGKAKLSGV 120
DB 61 TVIVFENHREKEJAVRVLKMGAKRVYVYHIIPAIADLKVRDLVLISGLTGGKAKLSGV 120

QY 121 RFIQEDYKVTVSALBGLDESAAQVMATYVNNLGYDGSITIGITDGDASHDPDLOGKV 180
DB 121 RFIQEDYKVTVSALBGLDESAAQVMATYVNNLGYDGSITIGITDGDASHDPDLOGKV 180

QY 181 IGWDFVNGRSYPTDGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADSGS 240
DB 181 IGWDFVNGRSYPTDGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADSGS 240

QY 241 ISTIIKGVENAVDNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNADAGIIVVCAAGNS 300
DB 241 ISTIIKGVENAVDNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNADAGIIVVCAAGNS 300

QY 301 GPNYTVGSPAAASKVITVGAJDSNDNLASFSGPTADGRKPEVAPGVDDIAPRSG 360
DB 301 GPNYTVGSPAAASKVITVGAJDSNDNLASFSGPTADGRKPEVAPGVDDIAPRSG 360

QY 361 TSMGPIINDYTKASGTSMAEYVSGVAGLILQAHPSTWPKVKTALIEDIAPKEIA 420
DB 361 TSMGPIINDYTKASGTSMAEYVSGVAGLILQAHPSTWPKVKTALIEDIAPKEIA 420

QY 421 DIAYGAGRVNYKAIKDYDDYAKLFTFGSVADKGSATHTFVSGATEVATLYWDTGSSDI 480
DB 421 DIAYGAGRVNYKAIKDYDDYAKLFTFGSVADKGSATHTFVSGATEVATLYWDTGSSDI 480

QY 481 DLYLYDPNGNEVDYSYATYGFKEKGYINPTAGTWTYKVSYGAAQYQVVDVSGLSQ 540
DB 481 DLYLYDPNGNEVDYSYATYGFKEKGYINPTAGTWTYKVSYGAAQYQVVDVSGLSQ 540

QY 541 SGGGNPNPNPNPTPTDQRTFGSVNDYWDTSFTFMVNSGAKITGDLTFTSYND 600
DB 541 SGGGNPNPNPNPTPTDQRTFGSVNDYWDTSFTFMVNSGAKITGDLTFTSYND 600

QY 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTLVYATYSTYCWADYQLKAVYVIG 559
DB 601 LDLYLYDPNGNLVDRSTSSNSYEHVEYANPAPGTWTLVYATYSTYCWADYQLKAVYVIG 559

RESULT 2

US-08-894-818B-35
Sequence 35, Application US/08094818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozi
APPLICANT: TSUNASAKA, Susumu
APPLICANT: KATO, Ikunoshi
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Broadway and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/98/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 332385/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-35

Query Match 87.48; Score 2997.5; DB 4; Length 654;
Best Local Similarity 88.38; Pred. No. 2.5e-210;
Matches 582; Conservative 12; Mismatches 50; Indels 5; Gaps 1;

QY 1 MKGLKALILVILVGLVGSVAAAPKKEQVNRVKNKGLTGLPKFKIQLKPNNEIS 60
DB 1 MKGLKALILVILVGLVGSVAAAPKKEQVNRVKNKGLTGLPKFKIQLKPNNEIS 60

QY 61 TVIVFENHREKEJAVRVLKMGAKRVYVYHIIPAIADLKVRDLVLISGLTGGKAKLSGV 120
DB 61 TVIVFENHREKEJAVRVLKMGAKRVYVYHIIPAIADLKVRDLVLISGLTGGKAKLSGV 120

QY 121 RFIQEDYKVTVSALBGLDESAAQVMATYVNNLGYDGSITIGITDGDASHDPDLOGKV 180
DB 121 RFIQEDYKVTVSALBGLDESAAQVMATYVNNLGYDGSITIGITDGDASHDPDLOGKV 180

QY 181 IGWDFVNGRSYPTDGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADSGS 240
DB 181 IGWDFVNGRSYPTDGHGTHVASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADSGS 240

QY 241 ISTIIKGVENAVDNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNADAGIIVVCAAGNS 300
DB 241 ISTIIKGVENAVDNKDKYGIKVINLSLGSQSSDGTDSLSQAVNNADAGIIVVCAAGNS 300

QY 301 GPNYTVGSPAAASKVITVGAJDSNDNLASFSGPTADGRKPEVAPGVDDIAPRSG 360
DB 301 GPNYTVGSPAAASKVITVGAJDSNDNLASFSGPTADGRKPEVAPGVDDIAPRSG 360

QY 361 TSMGPIINDYTKASGTSMAEYVSGVAGLILQAHPSTWPKVKTALIEDIAPKEIA 420
DB 361 TSMGPIINDYTKASGTSMAEYVSGVAGLILQAHPSTWPKVKTALIEDIAPKEIA 420

QY 421 DIAYGAGRVNYKAIKDYDDYAKLFTFGSVADKGSATHTFVSGATEVATLYWDTGSSDI 480
DB 421 DIAYGAGRVNYKAIKDYDDYAKLFTFGSVADKGSATHTFVSGATEVATLYWDTGSSDI 480

QY 481 DLYLYDPNGNEVDYSYATYGFKEKGYINPTAGTWTYKVSYGAAQYQVVDVSGLSQ 540
DB 481 DLYLYDPNGNEVDYSYATYGFKEKGYINPTAGTWTYKVSYGAAQYQVVDVSGLSQ 540

QY 541 SGGGNPNPNPNPTPTDQRTFGSVNDYWDTSFTFMVNSGAKITGDLTFTSYND 600
DB 541 SGGGNPNPNPNPTPTDQRTFGSVNDYWDTSFTFMVNSGAKITGDLTFTSYND 600

QY 601 LDLYLDPNGNLVDSTSSNSYEHVEYANPAGTWTFLVYAYSTYGHADYQKAVYVYG 659
DB 596 LDLYLDPNOKLVDRSESPNSYEHVEYLTLPAGTWTFLVYAYSTYGHAYTELAKYVYG 654

RESULT 3

US-09-445-472-16
; Sequence 16, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Nio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445.472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-445-472-16

Query Match 87.4%; Score 2997.5; DB 4; Length 654;
Best Local Similarity 88.3%; Pred. No. 2.5e-210;
Matches 582; Conservative 22; Mismatches 50; Indels 5; Gaps 1;

QY 1 MGLKALILVILVGLVGVGAAPEKVEQVRNVEKNYGLLTPGLFRKIOKLPNEIS 60
DB 1 MGLKALILVILVGLVGVGAAPEKVEQVRNVEKNYGLLTPGLFRKIOKLPNEIS 60
QY 61 TVIVFENHREKEIAVRVLELGMKAVRYVYHIIPAAADLKVRLVLSLGTGGKAKLSGV 120
DB 61 TVIVFENHREKEIAVRVLELGMKAVRYVYHIIPAAADLKVRLVLSLGTGGKAKLSGV 120
QY 121 RTIOEDKYVTVSAELEGDESAAQVNAVYVNNLCYDGSGITIGIDTGDASHDPDLOGKV 180
DB 121 RTIOEDKYVTVSAELEGDESAAQVNAVYVNNLCYDGSGITIGIDTGDASHDPDLOGKV 180
QY 181 ICWDFVNGRSYPYDONGHGTWASTAAGTGAASNGYKGNAPGAKLAGIKVLCADSGS 240
DB 181 ICWDFVNGRSYPYDONGHGTWASTAAGTGAASNGYKGNAPGAKLAGIKVLCADSGS 240
QY 241 ISTIKGVEMAVDNKDKYGIKVINLSLGSOSDGTDLSCAVNNAWDAGTVVCVAGHS 300
DB 241 ISTIKGVEMAVDNKDKYGIKVINLSLGSOSDGTDLSCAVNNAWDAGTVVCVAGHS 300
QY 301 GPNTVTVCSAAASKVITVGVNDNDNTAFSSRGPTADGRKLPVAPGVDTIAPRAGS 360
DB 301 GPNTVTVCSAAASKVITVGVNDNDNTAFSSRGPTADGRKLPVAPGVDTIAPRAGS 360
QY 361 TSMGPTINDYITKASCTSNATPRVSCVLCALIQHPSWTPDKVKTLTETADIVAPKEIA 420
DB 361 TSMGPTINDYITKASCTSNATPRVAGIAALLQAHPSWTPDKVKTLTETADIVAPKEIA 420
QY 421 DIAYGAGRVNAVYKAIKDYAKLTFTGSAVDKGSATHTFDVSGATFYATLYWDTGSSDI 480
DB 421 DIAYGAGRVNAVYKAIKDYAKLTFTGSAVDKGSATHTFDVSGATFYATLYWDTGSSDI 480
QY 481 DLVLYDPNGNEVDYSTAYTGEKVGYYNPNTAGTWVVKVYSGAANTQVDVSDGSL90 540
DB 481 DLVLYDPNGNEVDYSTAYTGEKVGYYNPNTAGTWVVKVYSGAANTQVDVSDGSL90 540
QY 541 SCGGNPNPNPNPTFTDTOTFTGSDVNDYDTSFTFNNVNSGATKLTGDLTFTDYSND 600
DB 541 -----PGSSPSPPQETTVDAKTFQSGDHYHYDRSDTFTFNVASGATKLTGDLTFTDYSND 600

QY 601 LDLYLDPNGNLVDSTSSNSYEHVEYANPAGTWTFLVYAYSTYGHADYQKAVYVYG 659
DB 596 LDLYLDPNOKLVDRSESPNSYEHVEYLTLPAGTWTFLVYAYSTYGHAYTELAKYVYG 654

RESULT 4

US-08-894-8188-1
; Sequence 1, Application US/088948188
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Nio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Shigumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894.8188
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-894-8188-1

Query Match 85.0%; Score 2914; DB 4; Length 659;
Best Local Similarity 83.2%; Pred. No. 3e-204;
Matches 538; Conservative 47; Mismatches 46; Indels 16; Gaps 7;

QY 1 MKRLKALILVILVGLVGVGAAPEKVEQVRN---VEKNYGLLTPGLFRKIOKLPNE 57
DB 1 MKRLGAVLALVGLVGLAGTALAAPKVPV--VRNNAVQKNYGLLTPGLFRKIOKLPNE 58
QY 58 EISTVIVENHREKEIAVRVLELGMKAVRYVYHIIPAAADLKVRLVLSLGT--TC--G 113
DB 59 EVDVTINFGSGDRRAVKRLKLMGAQVYKSTIIPAVAVKIKARDLLIAGMIDTGYFG 118
QY 114 KAKLSGVRFTIDEDYKTVTVSAELEGDESAAQVNAVYVNN-LGYDGSGITIGIDTGDAS 172
DB 119 NTRVSGIKFTIODEYKQVQVDA-----TSVSGIADTVNNSLGYDGSVVVAIVDTGDAN 173

Wed Nov 6 14:29:41 2002

114 KAKLSGVBFOEDKVTYSAZLEGLDESAAOVNATYVNN-LGYDGSGITIGITIGIDAS 172
119 NTRVSGIIFQEDIKVQVDDA-----TSVSGIGADTYNLSIGYDGSVVVAIVDTGIDAN 173
173 HPDLQKGVIGWDFVNGRSYPTDDHGHGTHVASTAAGTGAASNGKYKMGACAKLAGIKV 232
174 HPDLKGVIGWYDAVNGRSTPTDDHGHGTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVK 232
233 LGADSGSISTITIKGVENAYDNKDKYIKVNLGLSGSSQSDGTSLSQAVNNANDAGIV 292
233 LGADSGSVSTIAGDMVYQNKDKYIRVNLGLSGSSQSDGTSLSQAVNNANDAGIV 292
293 VCVAAAGSGPNITVSGPAAASVITVGAVDSDNDNIASFSSRGPTADGRKPEVYAPGV 352
293 VCVAAAGSGPNITVSGPAAASVITVGAVDSDNDNIASFSSRGPTADGRKPEVYAPGV 352
353 IAPRASGTSMTPTINDYTKASGTSMATPHVSGVGLILQAHPSWTPDKVKTALETAD 412
353 IAPRASGTSMTPTINDYTKASGTSMATPHVSGVGLILQAHPSWTPDKVKTALETAD 412
413 IVAPKEIADIAGAGRVNFKAYDDYAKLFTGTVADKGSATHTFDVSGATFVTATLY 472
413 IVAPKEIADIAGAGRVNFKAYDDYAKLFTGTVADKGSATHTFDVSGATFVTATLY 472
473 WDTGSSDIDLXYDPNGNEVDYSTAYTGFEKVGYYNPTAGTWIVKVVSYKGAANYQDV 532
473 WDTGSSDIDLXYDPNGNEVDYSTAYTGFEKVGYYNPTAGTWIVKVVSYKGAANYQDV 532
533 VSDGSLSSQSGGNPNPNPPTPTDQTFGTVADKGSATHTFDVSGATFVTATLY 592
533 VSDGSLSSQSGGNPNPNPPTPTDQTFGTVADKGSATHTFDVSGATFVTATLY 592
593 TFDTSYNDLXYDPNGNLDVRSSTSSYEHVEYANPAGTWTFLVYAYSTYGMADYQL 652
593 TFDTSYNDLXYDPNGNLDVRSSTSSYEHVEYANPAGTWTFLVYAYSTYGMADYQL 652
653 KAVVYIG 659.
653 KAVVYIG 659

RESULT 6
US-08-894-818B-3
Sequence 3, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, MIO
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoskin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Broadway and Melmark
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894.818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

173 HPDLQKGVIGWDFVNGRSYPTDDHGHGTHVASTAAGTGAASNGKYKMGACAKLAGIKV 232
174 HPDLKGVIGWYDAVNGRSTPTDDHGHGTHVAGIVAGTGSV-NSQYIGVAPGAKLVGVK 232
233 LGADSGSISTITIKGVENAYDNKDKYIKVNLGLSGSSQSDGTSLSQAVNNANDAGIV 292
233 LGADSGSVSTIAGDMVYQNKDKYIRVNLGLSGSSQSDGTSLSQAVNNANDAGIV 292
293 VCVAAAGSGPNITVSGPAAASVITVGAVDSDNDNIASFSSRGPTADGRKPEVYAPGV 352
293 VCVAAAGSGPNITVSGPAAASVITVGAVDSDNDNIASFSSRGPTADGRKPEVYAPGV 352
353 IAPRASGTSMTPTINDYTKASGTSMATPHVSGVGLILQAHPSWTPDKVKTALETAD 412
353 IAPRASGTSMTPTINDYTKASGTSMATPHVSGVGLILQAHPSWTPDKVKTALETAD 412
413 IVAPKEIADIAGAGRVNFKAYDDYAKLFTGTVADKGSATHTFDVSGATFVTATLY 472
413 IVAPKEIADIAGAGRVNFKAYDDYAKLFTGTVADKGSATHTFDVSGATFVTATLY 472
473 WDTGSSDIDLXYDPNGNEVDYSTAYTGFEKVGYYNPTAGTWIVKVVSYKGAANYQDV 532
473 WDTGSSDIDLXYDPNGNEVDYSTAYTGFEKVGYYNPTAGTWIVKVVSYKGAANYQDV 532
533 VSDGSLSSQSGGNPNPNPPTPTDQTFGTVADKGSATHTFDVSGATFVTATLY 592
533 VSDGSLSSQSGGNPNPNPPTPTDQTFGTVADKGSATHTFDVSGATFVTATLY 592
593 TFDTSYNDLXYDPNGNLDVRSSTSSYEHVEYANPAGTWTFLVYAYSTYGMADYQL 652
593 TFDTSYNDLXYDPNGNLDVRSSTSSYEHVEYANPAGTWTFLVYAYSTYGMADYQL 652
653 KAVVYIG 659
653 KAVVYIG 659

RESULT 5
US-09-445-472-12
Sequence 12, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, MIO
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoskin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472
CURRENT FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 12
LENGTH: 659
TYPE: PRT
ORGANISM: Thermococcus celer
US-09-445-472-12

Query Match 85.0%; Score 2914; DB 4; Length 659;
Best Local Similarity 83.7%; Pred. No. 3e-204;
Matches 558; Conservative 47; Mismatches 46; Indels 16; Gaps 7;
QY 1 MGLKALILVLYGLVYGVAAAEKKVEQVRN---VBKNYGLLTGCLFERRKQKLNPN 57
DB 1 MRUGAVVLAIVLGLAGTALAAAPKVP--VNNNAVQKNYGLLTGCLFERRKQKLNPN 58
QY 58 EISTVIFENHREKTAIVRLKMGAKVRYVYHIIPAIADLKVRDLIVISGL--TG--G 113
DB 59 EVDTVNFGSYGDRDRAVKVLRMGAAQVKYKIIPAVAVKIKARDLLLIAGHIDYGYG 118

APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 13-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 522 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: /note= xaa at position 428 is Gly or Val.
US-08-894-818B-3

Query Match 68.5%; Score 2349.5; DB 4; Length 522;
Best Local Similarity 85.2%; Pred. No. 2.7e-163;
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELEGDESAQVMAATYNNLYGDSGTTIGITGIDASHPDLOGKVIQWDFVNGRSY 192
DB 1 AELEGDESAQVMAATYNNLYGDSGTTIGITGIDASHPDLOGKVIQWDFVNGRSY 60
QY 193 PYDHGCHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAV 252
DB 61 PYDHGCHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAV 120
QY 253 DNKDKYGIKVINLSLSSOSSDGTSLQAVNNAWDAGIIVCVAAGNSGPNITYTIGSPAA 312
DB 121 DNKDKYGIKVINLSLSSOSSDGTSLQAVNNAWDAGIIVCVAAGNSGPNITYTIGSPAA 180
QY 313 ASKVTYVGVDSNDINAFSSRGPTADGRKLPKVPWAPGVVDIIAPRASCTSNGTINDYIT 372
DB 181 ASKVTYVGVDSNDINAFSSRGPTADGRKLPKVPWAPGVVDIIAPRASCTSNGTINDYIT 240
QY 373 KASCTSNATPHVSGVALLQAHPSWTPDKVKTALITADIAPKEIADIAGAGRVNY 432
DB 241 AAPTSTNATPHVSGVALLQAHPSWTPDKVKTALITADIAPKEIADIAGAGRVNY 300
QY 433 KAIKYDDYAKLTFTGSVADKGSATHTFDVSCATFVTATLYMDTSSDIDLKLYDPNGNEV 492
DB 301 KAINVDYAKLVFTGVANKGSQTHQFVIGSASFVTATLYMDNANSLDLKLYDPNGNOV 360
QY 493 DYSTATYGFKEKVGYYNPTACTWTWKVYKGAANYOVVDVSDGSLSSGCGNPNPNP 552
DB 361 DYSTATYGFKEKVGYYNPTACTWTWKVYKGSANTOVVDVSDGSLSSGCGNPNPNP 415
QY 553 NPTFTDQFTGSDVNDYDTSFTTMVNSGATKITGDLTSDTSDNLDLYDPNGNL 612
DB 416 QPEPTVDKATFOXSDHYDRSDTFTMTVNSGATKITGDLVFTDTSYHDLKLYDPNQKL 475
QY 613 VDRSTSNSEYHEVYANAPGTWFLVYAVSTYGMADYOLKAVVYVG 659
DB 476 VDRSESPNSEYHEVYLTAPGTWFLVYATYTCNAYIELTAKVYIG 522

RESULT 7
US-09-445-472-4
Sequence 4, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, HIKARU
APPLICANT: MORISHITA, MIO
APPLICANT: SHIMOJO, TOMOKO
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472

APPLICANT: KATO, IKUOSHI
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 522
TYPE: PRT
ORGANISM: Pyrococcus furiosus
FEATURES:
NAME/KEY: misc_feature
LOCATION: (428)..(428)
OTHER INFORMATION: xaa at position 428 is Gly or Val.
US-09-445-472-4

Query Match 68.5%; Score 2349.5; DB 4; Length 522;
Best Local Similarity 85.2%; Pred. No. 2.7e-163;
Matches 449; Conservative 22; Mismatches 51; Indels 5; Gaps 1;

QY 133 AELEGDESAQVMAATYNNLYGDSGTTIGITGIDASHPDLOGKVIQWDFVNGRSY 192
DB 1 AELEGDESAQVMAATYNNLYGDSGTTIGITGIDASHPDLOGKVIQWDFVNGRSY 60
QY 193 PYDHGCHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAV 252
DB 61 PYDHGCHGTHVASIAAGTGAASNGKYGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAV 120
QY 253 DNKDKYGIKVINLSLSSOSSDGTSLQAVNNAWDAGIIVCVAAGNSGPNITYTIGSPAA 312
DB 121 DNKDKYGIKVINLSLSSOSSDGTSLQAVNNAWDAGIIVCVAAGNSGPNITYTIGSPAA 180
QY 313 ASKVTYVGVDSNDINAFSSRGPTADGRKLPKVPWAPGVVDIIAPRASCTSNGTINDYIT 372
DB 181 ASKVTYVGVDSNDINAFSSRGPTADGRKLPKVPWAPGVVDIIAPRASCTSNGTINDYIT 240
QY 373 KASCTSNATPHVSGVALLQAHPSWTPDKVKTALITADIAPKEIADIAGAGRVNY 432
DB 241 AAPTSTNATPHVSGVALLQAHPSWTPDKVKTALITADIAPKEIADIAGAGRVNY 300
QY 433 KAIKYDDYAKLTFTGSVADKGSATHTFDVSCATFVTATLYMDTSSDIDLKLYDPNGNEV 492
DB 301 KAINVDYAKLVFTGVANKGSQTHQFVIGSASFVTATLYMDNANSLDLKLYDPNGNOV 360
QY 493 DYSTATYGFKEKVGYYNPTACTWTWKVYKGAANYOVVDVSDGSLSSGCGNPNPNP 552
DB 361 DYSTATYGFKEKVGYYNPTACTWTWKVYKGSANTOVVDVSDGSLSSGCGNPNPNP 415
QY 553 NPTFTDQFTGSDVNDYDTSFTTMVNSGATKITGDLTSDTSDNLDLYDPNGNL 612
DB 416 QPEPTVDKATFOXSDHYDRSDTFTMTVNSGATKITGDLVFTDTSYHDLKLYDPNQKL 475
QY 613 VDRSTSNSEYHEVYANAPGTWFLVYAVSTYGMADYOLKAVVYVG 659
DB 476 VDRSESPNSEYHEVYLTAPGTWFLVYATYTCNAYIELTAKVYIG 522

RESULT 8
US-09-445-472-1
Sequence 1, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, HIKARU
APPLICANT: MORISHITA, MIO
APPLICANT: SHIMOJO, TOMOKO
APPLICANT: ASADA, KIYOZO
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472

;; CURRENT FILING DATE: 1999-12-06
;; PRIOR APPLICATION NUMBER: 151969/1997
;; PRIOR FILING DATE: 1997-06-10
;; NUMBER OF SEQ ID NOS: 33
;; SOFTWARE: Patent in version 3.0
;; SEQ ID NO 1
;; LENGTH: 412
;; TYPE: PRT
;; ORGANISM: Pyrococcus furiosus
US-09-445-472-1

Query Match 56.0%; Score 1918; DB 4; Length 412;
Best Local Similarity 89.5%; Pred. No. 5.1e-132;
Matches 367; Conservative 15; Mismatches 28; Indels 0; Gaps 0;
QY 133 AELEGLESAAOVMATYVNNLCYDGGTITIGTIDTGDASHPDLOGKVGIVGVNGRSY 192
DB 1 AELEGLESAAOVMATYVNNLCYDGGTITIGTIDTGDASHPDLOGKVGIVGVNGRSY 60
QY 193 PYDRHGHTIVASIAAGTGAASNGKYKCHAPGAKLAGIKVLCADGSGSISFLIKGVEMAY 252
DB 61 PYDRHGHTIVASIAAGTGAASNGKYKCHAPGAKLAGIKVLCADGSGSISFLIKGVEMAY 120
QY 253 DNKDKYGIKVINLSLSSQSSDGTDSLQAVNNAMADAGIWCVAAGNSGPNNTYVGSPPAA 312
DB 121 DNKDKYGIKVINLSLSSQSSDGTDSLQAVNNAMADAGIWCVAAGNSGPNNTYVGSPPAA 180
QY 313 ASKVITVGVNDNDNTASFSRGPADGRKPEVWAPGVVDIAPRASGTSMTGTPINDYIT 372
DB 181 ASKVITVGVNDNDNTASFSRGPADGRKPEVWAPGVVDIAPRASGTSMTGTPINDYIT 240
QY 373 KASGTSNATPHVSGVAGALILQAHPSWTPDKVKTALITADIYAPKEIADIYAGRVNY 432
DB 241 AAPTGSNATPHVAGIALLQAHPSWTPDKVKTALITADIYAPKEIADIYAGRVNY 300
QY 433 KAIKYDDYAKLITFTSGVADGSAATHFDVSGATFTVATLYWDGSDIDLILYDPNGEY 492
DB 301 KALINDTAKLVFTCYVANKGSGTHQFVLSGASFTVATLWDNANSDDLILYDPNGNOV 360
QY 493 DYSYATYGFKEKGVYVNPAGTWTYKVSYSYGAANYOVVDVDSLSQSG 542
DB 361 DYSYATYGFKEKGVYVNPAGTWTYKVSYSYGAANYOVVDVDSLSQSG 410

RESULT 9
US-08-750-532-18
; Sequence 18, Application US/08750532
; Patent No. 5756339
; GENERAL INFORMATION:
; APPLICANT: MITTA, Masanori
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MORISHITA, Mio
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY, ROBERT, P.L.L.C.
; STREET: 419 Seventh Street N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,532
; FILING DATE: 1998-12-06
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/01095
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1994/130236
; FILING DATE: 13-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1994/173912
; FILING DATE: 26-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, ROGER L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: MITTA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-750-532-18

Query Match 30.5%; Score 1045.5; DB 1; Length 237;
Best Local Similarity 87.4%; Pred. No. 8e-69;
Matches 202; Conservative 9; Mismatches 19; Indels 1; Gaps 1;
QY 175 DLQKVKIVGVNGRSYVDDHGHTVASTAAGTGAASNGKYKCHAPGAKLAGIKVLC 234
DB 1 DLQKVKIVGVNGRSYVDDHGHTVASTAAGTGAASNGKYKCHAPGAKLAGIKVLC 59
QY 235 ADGSGSSTRIKGVEMAYVNDKYGKIVNLSLSSQSSDGTDSLQAVNNAMADAGIYVC 294
DB 60 ADGSGSSTRIKGVEMAYVNDKYGKIVNLSLSSQSSDGTDSLQAVNNAMADAGIYVC 119
QY 295 VAAGNSGPNNTYVGSPPAAKSVITVGVNDNDNTASFSRGPADGRKPEVWAPGVVDII 354
DB 120 VAAGNSGPNNTYVGSPPAAKSVITVGVNDNDNTASFSRGPADGRKPEVWAPGVVDII 179
QY 355 APRASGTSMTGTPINDYITKASGTSNATPHVSGVAGALILQAHPSWTPDKVKT 405
DB 180 APRASGTSMTGTPINDYITKASGTSNATPHVSGVAGALILQAHPSWTPDKVKT 230

RESULT 10
US-09-000-016-4
; Sequence 4, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.5 inch, 1.44 mb
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER:
:
: FILING DATE:
:
: ATTORNEY/AGENT INFORMATION:
:
:   NAME: Warren M. Check, Jr.
:   REGISTRATION NUMBER: 33,367
:
: REFERENCE/DOCKET NUMBER:
:
: TELECOMMUNICATION INFORMATION:
:
:   TELEPHONE: 202-721-8200
:   TELEFAX: 202-721-8250
:
:   TELEX:
:
: INFORMATION FOR SEQ ID NO: 4:
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 734 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: protein
:
: US-09-000-016-4

```

Query Match	20.28;	Score 692.5;	DB 4;	Length 734;
Best Local Similarity	33.60;	Pred. NO. 2e+42;		
Matches 202;	Conservative 62;	Mismatches 223;	Indels 115;	Gaps 21;

Qy	110	LTGCKAKLCSVRPFIOEDYKVTYSAELEGJLDESNAQVWATYWNLYGVDSGTIGTIDTGI	169
Db	182	VTNGDRTFASGIAHVLD-----GVRRAALJDTSVGOIGAPKAWSGDYDKGKYTAVDTDV	236
Qy	170	DASHPDLOQKVIQWYDFYNGRSTPYDDHGRGTHYASIAAGTGAASNGKYKGNAPKAGLAK	229
Db	237	DTSHPDLKGRVYASKNFYAAPG-AGDKVCHGTHYASIAAGTGAOSKGYKCVAPGAALIN	295
Qy	230	IKVLGADGSGSITIIKGVENKAVDNDKYGINVNLISLGSOSSDGTJDSLSQAVNN-AND	288
Db	296	GRVLDDSGFDGSLGACNEKA-----AGADAVNMSLGGKNDTPE-TDPLEAAVDKLSAE	350
Qy	289	AGIVCYAAGNSGPNITYTGPSAAASKYITVGAVDSNDNIASFSSRGP-TADGRKPEYV	347
Db	351	NGVLFATAAGNEQPE-SIGSPGSADAALTYGAVDDKDLADFSTSPRLGDGAIKPDVT	408
Qy	348	APGVDDIAPRASGTSKGTPIIND---YVTKASGTSWATPHYSGVGLILQAHPSFTDKVK	404
Db	409	APGVDTTASASGNDIGQVEGEGPAGYWTISGTSWATPHVAGAAALLKQHPDWTSELK	468
Qy	405	TALLETADIVAFREIADIYAGAGRVNYYKAIK-----	436
Db	469	GAL--TGSTKGGK-YTFPEOGSGRIQADKALQOATVIADPVSVSGVQOHPHTDDEPVTKO	525
Qy	437	-----YDYAKLFTFGSVADKGSATHTFDVSGATFTVATLYMDTSGSDIDILYLDPN	488
Db	526	LTYNLGTQDVTILKLTSTATDPKGAAPAGFTTGGATTVPYA---GGSASVDWTADTRL	582
Qy	489	CNEVDYSTYAIYGFENKYGNPTAGTWTYKVVSY-----KGAANYCVDVDSGDS	537
Db	583	GGTVDGAYSAYVYVATGGGTGYRTAAAVQREVEYSQVTVRIHGRDGRKPTTEHLTDLIGYAG	642
Qy	538	LSQSGGNGPNPNPNPTPTTDTQTF-----TGS-VANDYWDTSDFTFW-----NVNSEA	585
Db	643	LCSCRGYV-----APATDTATRLPKGYTLVDSWIAKDFGLKGDMLVQPKLSV	693
Qy	586	TKLITGDLTFTD-SYNDLILYLDPNGLNVDRTSSNSYEHVEYANPACPTWTFLYYAST	644
Db	694	TKDT-TLTLDARTTKAADITVDDPK-----AKPLSAT---IGYTYDT	731
Qy	645	YG	646
Db	732	AG	733

RESULT 11

US-09-000-016-2 ; Sequence 2, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:

QY 489 GNEVDYVYATYGEKVGYNPNAGTWTKVVS-----KGAANYQVVDVSGS 537
DB 583 GGTVDGAYVATYGGQVTRTAAVQREVSVDVTVRHIGRDKPTTEHLTDLIGVAG 642
QY 538 LSCSGGNPNPNPTPTDTDTQTF---TGS-VNDYVDTSDTFTM-----NVNSGA 585
DB 643 LSGRGY-----APATDTATLRLPKCTYVDSHIAKDFCTLAGGIDWLVPKLSV 693
QY 586 TKITGDLPT-SYNDLGLYDENGKLVDRSTSSNHYEYANPAGTFLVYAYST 644
DB 694 TKDT-TLTDARTTAAADITVPOFK-----AKPLSAT---IGTYDT 731
QY 645 YG 646
DB 732 AG 733

RESULT 12

US-09-000-016-7
; Sequence 7, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 Inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 520 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-000-016-7

Query Match 19.8%; Score 679.5; DB 4; Length 520;
Best Local Similarity 36.3%; Pred. No. 1.1e-41;
Matches 181; Conservative 55; Mismatches 183; Indels 79; Gaps 15;

QY 138 LQSAQVNAVYVWNLGVDGSGITIGIDGIDASHPDLOGKVGWVDFVNGRSYPDDH 197
DB 1 LQSVGIGAPKAWAGYDGKVIATLDGVDTSHPDLKGRVATSKNFTAAPG-AGKAV 50
QY 198 GHGTHVASTAGTGAASNGKYKGNAPGAKLAGIKVLGADGSGSITIKGVENAVDKK 237

DB 60 GHGTHVASTAGTGAQSGKYGKVPAGAAIILGKVLDDSGGDDSGILAGNEHAA-----A 115
QY 258 YGIKVINLSGSSQSDGTCSLQAVNN-AWDAGIIVCVAAAGNSGPNYTVVGSPPAAASKV 316
DB 116 QGADVNNLSIGMDTPE-TDPLEAAVDKLSAEKGVLFALAGNEGPE--SIGSPGSADAA 172
QY 317 ITVGAVDSNDNTASFSRGP-TADGRILKPEVAPGVDDIIPRASGTSMTGTPIND---YVT 372
DB 173 LTVGAVDDKDLADSSFGRLGCAIKPDVTAGVYDITAAAEAGNDIGQEVGSGPAGYM 232
QY 373 KASGTSMATHPSVGVGALILQAHSPWPKVKPTALITETADIIVAPKEIADIAYGAGRNVY 432
DB 233 TISGTSNATHPVAGAAALLKQCHPDMTSABLKAL--TGSTRGK-YPTEPQSGSGRIQAD 289
QY 433 KAIK-----YDDYAKLFTFTGTVADKGSAT 456
DB 290 KALQCTVIADPVSVSGVQOQPHDTDEPTVKLTLYRLNLTQDVTTLKLTSTATDPKRAAP 349
QY 457 HTFDVSGATFVATLYWDTGSSDIDLILYDPNGNEVDYSTAYYGEKVGYNFTAGTWT 516
DB 350 AGFTLCATTVTVA---CGSA3VDMTADTRLGCTYDGSAYVAVATCGGQTVRTAAAVQ 406
QY 517 VKVSV-----KGAANYQVVDVSGSLSQSGGNGPNPNPTPTDTDTQTF-- 563
DB 407 REVESYDVTVRHICRGKPTTEHLTDLIGVAGLSGRGYG-----APATDTATLRL 457
QY 564 -TGS-VNDYVDTSDTFTM 579
DB 458 PKGYLVDSNIAKDFGTL 475

RESULT 13
US-08-685-774-2
; Sequence 2, Application US/08685774
; Patent No. 5712147
; GENERAL INFORMATION:
; APPLICANT: Shih, Jason C. H.
; APPLICANT: Lin, Xiang
; APPLICANT: Miller, Eric S.
; TITLE OF INVENTION: DNA ENCODING BACILLUS LICHENIFORMIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: NC 2712147th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,774
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US-08/250,028
; FILING DATE: 27-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 33,665
; REFERENCE/DOCKET NUMBER: 5051-260
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 420-2203
; TELEFAX: (919) 881-2175
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus licheniformis
STRAIN: PWD-1
US-08-685-774-2

Query Match 16.28; Score 557; DB 1; Length 379;

Best Local Similarity 35.08; Pred. No. 5.7e-33;

Matches 153; Conservative 62; Mismatches 148; Indels 74; Gaps 15;

Db 3 GLKALILVLVLGVLGVSAAPEKKVEQVRNVEKNYGLLTPGLFRKIOKLNPEEISIV 62
10 GMLTAPMLVFTWAFSDSASAQPAK-----NVEKDY-----I 41
Qy 63 IVFENH-REKETAVRLELGMCAKVRVYVHIIPAIADLKVRDLVLVLSGLTGGRAKLSGYR 121
Db 42 VPKSGVKTASYKVDVKEGKVDKOPRIINAAKAKLDEKALKEVK-----NDPDA 94
Qy 122 FIOEDYKVTVAELEGLEDSAAQVNAVYVNNLGYDGSITIGIIDTIDASHPDLOGKVI 181
Db 95 YVEEDH--VAHALAOTVPYGIPLIKADKVOAQCPKGANVYKVALDVTGIQASHEDL--NVV 150
Qy 182 CWDFVNGSYSPYDDHGHGTHVASYAAGTGAASNGK--YKMGAPGAKLAGIKVLGADSG 239
Db 151 GGASPVAGAYWTDGNGHGHV-----AGTVAALDNTGVLGVPVPSVSLYAVKVLNLSGSG 206
Qy 240 SISTIIGVEMADVNDKDKYIKVINLSGSSQSDGTSLDLSQAVNNAWDAGIYVCAAGN 299
Db 207 SYSGIVSGIEMATTN----GMDVINMSLG---GASGTAMQAVDNAVARGVYVVAAGN 259
Qy 300 SGP--NTYTVGSPAAASKVIITVGAVDSNDNIASFSSRGPTADGRLEKPEYVAGVDIAPR 357
Db 260 SGSGTNTIGYPAKYDSVIANGVDSNRSFSSVG-----AELEYMAPG----- 306
Qy 358 ASGTSNGTINDYITKASGTSMATPHVSGVGLIILQAHPSWTPDKYKALIEPIADIVAPK 417
Db 307 -AGVYSTYPTNTYAT-LNCTSHVSPRVAGAAALILSKHPNLSASQVRNLSSTATYLG-- 362
Qy 418 EIADIAYGAGRVNYKA 434
Db 363 --SSFYGKGLINVEAA 377

RESULT 14

US-09-445-472-31

Sequence 31, Application US/09445472

Patent No. 6358726

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikoaru

APPLICANT: NORISHITA, Mio

APPLICANT: SHIMOJO, Tomoko

APPLICANT: ASADA, Kiyozo

APPLICANT: KATO, Ikunoshi

TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA-6

CURRENT APPLICATION NUMBER: US/09/445,472

CURRENT FILING DATE: 1999-12-06

PRIOR APPLICATION NUMBER: 151969/1997

PRIOR FILING DATE: 1997-06-10

NUMBER OF SEQ ID NOS: 33

SOFTWARE: Patentin version 3.0

SEQ ID NO 31

LENGTH: 382

TYPE: PRT

ORGANISM: Bacillus subtilis

US-09-445-472-31

Query Match 16.28; Score 555; DB 4; Length 382;

Best Local Similarity 35.18; Pred. No. 8.1e-33;

Matches 156; Conservative 65; Mismatches 141; Indels 86; Gaps 18;

Qy 1 MKGLKALILVLVLGLV-----VGSVAAAPEKKVEQVRNVEKNYGLLTPGLFRKIQKLNPN 56

Db 1 MRGKKVITSLLEFALCALIFTWAPGSTSSA---QAACKSNGEKKY----- 40
Qy 57 BEISTVIVFENHREKETAVR---VLEUMCAKVRVYVHIIPAIADLK---VRDLLVISGL 110
Db 41 -----IVGEKQTMSTN-AARKKDVIBSEKGGYKQKQKYVDAASATLHEKAVKEL----- 89
Qy 111 TGGKAKLSGVREFIOEDYKVTVAELEGLEDSAAQVNAVYVNNLGYDGSITIGIIDTID 170
Db 90 ----KKDPSAVVEEDH--VAHAVAQSPVPYGVSOIKAPALHSQCYTCSNVKVAVIDSGID 143
Qy 171 ASHPDLOGKVICWDFVNGSYSPY-DDHGHGTHVASYAAGTGAASNGK--YKMGAPGAKL 227
Db 144 SSPDL--KVAGSGMWPSETNPFQDNNSRGTHV-----AGTVAALNNSIGVLGVAPASL 197
Qy 228 AGIKVLGADSGSTISTINGVEMADVNDKDKYIKVINLSGSSQSDGTSLDLSQAVNNAW 287
Db 198 YAVKVLGADSGSQYSWIINGIEHALAN---NNDVINMSLG---GPSGAALKAADVDAV 250
Qy 288 DAGIYVCAAGNSGP--NTYTVGSPAAASKVIITVGAVDSNDNIASFSSRGPTADGRLEKPE 345
Db 251 ASGVVVAAGNEGTSSTSVGPKYPSYAVGAVDSSNQSRASPSVSGPELD----- 304
Qy 346 VVAPGVDDIAPRASGTSNGTINDYITKASGTSMATPHVSGVGLIILQAHPSWTPDKYK 405
Db 305 VMAPGVSI-----OSTIACNKYKAYNGTSMASPHVAGAAALILSKHPNLTNTQVRS 355
Qy 406 ALIETADIVAPKEIAD-IAYGAGRVNYKA 434
Db 356 SLENTI-----TKLGDSPFYGKGLINVOA 380

RESULT 15

US-07-772-087-4

Sequence 4, Application US/07772087

Patent No. 5275945

GENERAL INFORMATION:

APPLICANT: HSIAO, Hung-Yu

APPLICANT: FODGE, Douglas W.

APPLICANT: LALONDE, James J.

TITLE OF INVENTION: ALKALINE PROTEASES STABLE IN HEAVY-DUTY

TITLE OF INVENTION: DETERGENT LIQUIDS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: Virginia

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: P-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/772,087

FILING DATE: 19911003

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 25,768

REFERENCE/DOCKET NUMBER: 16754/115 CHCO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 378 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

us-09-841-553-5.ra1

Wed Nov 6 14:29:41 2002

```

: ORIGINAL SOURCE:
: ORGANISM: Bacillus
: STRAIN: Carlsberg
: US-07-772-087-4

Query Match      16.2a; Score 554.5; DB 1; Length 378;
Best Local Similarity 34.2a; Pred. NO. 8.7e-33;
Matches 149; Conservative 64; Mismatches 150; Indels 73; Gaps 13;

QY 3 GLKALILVILGLVGSVAAAPKKVEQVRNVEKNVGLTTPGLFRKIOKLNPNEEIETV 62
Db 10 GMLTAPMLVFTMAESDSASAAQPAK-----NVERDY-----I 41
QY 63 IVEFNHREKEIAVRLEMLGAKVRYVYHITPAIADIKVRLDLYISLIGGKAKLSGVRF 122
Db 42 VGEKSGVKTASYKKDKIKESGKVDKQFRIINAAKAKLDKEALKEVK-----NDPDVAY 94
QY 123 IQEDYKVTVSAREGLEDSRAAQVNATYVNNLSYDGSGITIGITGDASHPDLOGKVG 182
Db 95 VEEDH--VAHALAQTPVYGIPLIKADKVOAQGFKANVAVLDTGTQASHPD--NVYG 150
QY 183 WYDFYNGRSYPYDDHGHGTHVASTAGTGAASNGK--YKGMAPGAKLAGIKVLGADGGS 240
Db 151 GASFVAGERAYNTDGNHGHV----AGTVAALDNTTGVGVAPSVLYAYKVLNSSGSGT 206
QY 241 ISTIIKGVEMAYDNKDKYGIKVINLSLGSQSDGCTDLSQAVNNAMDAGIVVCVAAAGNS 300
Db 207 YSGIVSGIEMATTN---GNDVINMSLG---GPSGSTANKOAVDNAYARGVYVYVAAAGNS 259
QY 301 GP--NTYTVGSPAASKVITTCVAVDSNDNTASFSSHGPTADGRKLPKPEVAPGVDTIAPRA 358
Db 260 GSSGNTNTIGYPARKYDSVIAYGAVDSNNSRASFSSVG-----AELEVMAPGAGVYSTYP 313
QY 359 SGTSMCTPIINDYTKASGTSMATPHVSGVGALILQAHPSWTPDKVKYKLTATETADIAPKE 418
Db 314 TST-----YATLNGTSMASPHVAGAAALILSKHPNLSASQVRNRLSSTATYLG--- 361
QY 419 IADIYAGAGRVNYKA 434
Db 362 -SSFYKGLINVEAA 376

```

Search completed: October 31, 2002, 13:32:28
Job time : 15 secs